WO9109942

Publication Title:

SITE-SPECIFIC DOUBLE STRANDED DNA ENDONUCLEASE

Abstract:

Abstract of WO9109942

An endonuclease which cleaves double stranded DNA at a specific site, producing staggered ends with 4 base pair, 3' overhangs, has been isolated, purified and characterized. A method of producing sufficient quantities of the endonuclease for purification and characterization is disclosed. A method to use the endonuclease to cleave DNA, producing fragments useful for gene mapping is also disclosed. Data supplied from the esp@cenet database - Worldwide

---

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12N 9/22, 15/10	A2	(11) International Publication Number: WO 91/09942 (43) International Publication Date: 11 July 1991 (11.07.91)

(21) International Application Number: PCT/US90/07644

(22) International Filing Date: 24 December 1990 (24.12.90)

(30) Priority data:
456,100 22 December 1989 (22.12.89) US

(60) Parent Application or Grant
(63) Related by Continuation
US Not furnished (CIP)

(71) Applicant *(for all designated States except US)*: BOARD OF REGENTS, THE UNIVERSITY OF TEXAS SYSTEM [US/US]; 201 West 7th Street, Austin, TX 78701 (US).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)* : BUTOW, Ronald, A. [US/US]; 6021 Northaven Road, Dallas, TX 75230 (US).

(74) Agent: PARKER, David, L.; Arnold, White & Durkee, P.O. Box 4433, Houston, TX 77210 (US).

(81) Designated States: AT, AT (European patent), AU, BB, BE (European patent), BF (OAPI patent), BG, BJ (OAPI patent), BR, CA, CF (OAPI patent), CG (OAPI patent), CH, CH (European patent), CM (OAPI patent), DE, DE (European patent), DK, DK (European patent), ES, ES (European patent), FI, FR (European patent), GA (OAPI patent), GB, GB (European patent), GR (European patent), HU, IT (European patent), JP, KP, KR, LK, LU, LU (European patent), MC, MG, ML (OAPI patent), MR (OAPI patent), MW, NL, NL (European patent), NO, RO, SD, SE, SE (European patent), SN (OAPI patent), SU, TD (OAPI patent), TG (OAPI patent), US.

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: SITE-SPECIFIC DOUBLE STRANDED DNA ENDONUCLEASE

(57) Abstract

An endonuclease which cleaves double stranded DNA at a specific site, producing staggered ends with 4 base pair, 3′ overhangs, has been isolated, purified and characterized. A method of producing sufficient quantities of the endonuclease for purification and characterization is disclosed. A method to use the endonuclease to cleave DNA, producing fragments useful for gene mapping is also disclosed.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GN	Guinea	NL	Netherlands
BJ	Benin	GR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland			SN	Senegal
CI	Côte d'Ivoire	KR	Republic of Korea	SU	Soviet Union
CM	Cameroon	LI	Liechtenstein	TD	Chad
CS	Czechoslovakia	LK	Sri Lanka	TG	Togo
DE	Germany	LU	Luxembourg	US	United States of America
DK	Denmark	MC	Monaco		

SITE - SPECIFIC DOUBLE STRANDED DNA ENDONUCLEASE

This invention relates generally to a protein, more specifically to an endonuclease protein which has not previously been purified or characterized from its natural biological source. This endonuclease is novel and extremely useful because it cleaves double-stranded DNA at specific, infrequent sites, for which endonucleases were not previously available. The resulting fragments are of great value for human gene mapping because the cleavage site is a sequence ordinarily encountered in genomic DNA, and because cleavage by the endonuclease produces relatively larger fragments than characteristic of those produced by many previously available endonucleases. This invention also includes methods for purifying the endonuclease and for cleaving DNA by use of the endonuclease.

(A) Restriction Endonucleases

One of the essential tools molecular biologists use to delve deeper into the mysteries of life contained in the structure of DNA, the genetic material, is a molecular scissors called a restriction endonuclease. There are many such enzymes which are capable of cutting DNA at specific sites (see Lewin, 1987 for review).

Restriction enzymes (restriction endonucleases) recognize specific short sequences of DNA (usually unmethylated DNA) and cleave the duplex molecule, usually at the target recognition site, but sometimes elsewhere. In some instances, the recognition site is specific, but the cleavage site is located some distance away from the

recognition site and does not appear to be at any specific sequence.

"Duplex" refers to the double stranded composition of the DNA molecule. The cleavage induced by endonucleases is usually at specific sequences of approximately 4-6 base pairs. A base pair is a union of purines or pyrimidines in the DNA duplex. There are four such bases and they pair in specific unions: adenine with thymine, (A-T), guanine with cytosine (G-C).

Fragments generated by endonucleases are amenable for further analysis of their nucleotide composition. Variation in the fragment sizes obtained from the same chromosomal locations among individuals, is referred to as restriction fragment length polymorphism (RFLP).

Restriction endonucleases are essential components of methods used to construct maps of the genetic material, although not all such endonucleases are useful. Some of the problems limiting their use are that cleavage may be too frequent using a particular enzyme, producing pieces too small to be useful. Another problem is that the sites attacked may have nucleotide sequences that are so unusual that they are not likely to occur in vivo. Some enzymes only cleave artificially engineered sequences.

Restriction endonucleases are named by using three or four letter abbreviations identifying their origin, coupled with a letter and/or number designation which distinguish multiple enzymes of the same origin. An example of the nomenclature is EcoR1, one of the endonucleases derived from E. coli. Most of the endonucleases discovered initially were isolated from bacteria, in which they cleave DNA as part of the natural function of the cell. However, other organisms, for

example, yeast, can be used as a source of double-strand DNA cleaving endonucleases.

Isolation of many endonucleases occurred because the bacteria from which the endonucleases were derived were able to distinguish between the DNA native to the bacteria and any invading foreign DNA. One of the ways bacteria recognize foreign DNA is by the absence of methyl groups at appropriate base pair sites. The bacteria protects its own DNA from cleavage by its own endonucleases, by methylation of its own DNA bases at appropriate target sites. Successful attack on bacteria by foreign DNA, for example by viruses, may be due either to the fact that the virus DNA has the same pattern as the host DNA, or alternatively, that mutations have caused defects in the ability of the bacteria to produce an endonuclease or to attack the foreign DNA. Endonucleases isolated from bacteria are of two types, one which is only able to cleave DNA, and another in which both restriction and methylation activities are combined. Some restriction endonucleases introduce staggered cuts with overhangs while others generate blunt ends.

(B) Restriction Mapping

Gene maps give the location of specific genes (specific DNA nucleotide sequences) that encode the primary sequences of protein gene products relative to each other and also localize the genes on specific chromosomes of higher organisms. A map of DNA obtained by using endonucleases to map breakpoints is called a restriction site map and consists of a linear sequence of cleavage sites. This physical map is obtained by extracting DNA from the chromosomes in cells, breaking the extracted DNA at various points with endonucleases,

and determining the order of cleavage sites by analysis of the fragments.

Distances along the maps are measured directly in base pairs or, if distances are long, in megabase pairs. By comparing the sequences of DNA between relatively short distances, a DNA map is constructed in a stepwise fashion. A major goal of current research is to construct a map of the entire human genome. (The Human Genome Project, American Society of Human Genetics Symposium, Baltimore, Nov. 15, 1989.) Success in mapping human and animal genomes will require a selection of endonucleases which cleave at a large variety of sites which occur in the DNA of living organisms, not just in artificial sequences.

DNA fragments produced by the action of endonucleases are separated on the basis of size by agarose or polyacrylamide gel electrophoresis. An electric current is passed through the gel, causing the fragments to move down it at a rate depending on length; the smaller fragments move more rapidly. The result of this migration in a gel, is a series of bands each corresponding to a fragment of a particular size. Many different endonucleases are used for gene mapping, and large numbers of overlapping fragments are analyzed. Sequential cleavage using different endonucleases produces a series of larger fragments broken down into smaller fragments. A hierarchy is then constructed based on the fact that there is complete additivity of length of the fragments within the original starting fragment. For example, a fragment of 2,100 base pairs may be broken down into 200 and 1900 base pairs. (see Lewin, 1987 for review).

Construction of an entire gene map for a species, for example construction of the human gene map, is a

difficult and tedious task. The larger the number of endonucleases available for restriction mapping, the easier and more sophisticated the genetic map construction. In particular, many endonucleases are needed which cleave at a variety of specific sites and which produce fragments of different lengths. To appreciate the magnitude of the mapping problem it should be noted that an estimated 3 billion base pairs contained in 22 pairs of human chromosomes called autosomes plus two sex chromosomes, comprise the human genome.

Restriction maps represent advantages over older methods of mapping which identified a series of genetic sites because of the occurrence of DNA changes (mutations), because restriction maps can be obtained for any sequence of DNA. Their construction is not dependent upon the location of mutations, and no knowledge of the function of a particular sequence of DNA is required. However, restriction maps are related to, and are generally colinear with, "genetic" maps.

Mutations which are deletions or insertions of base pairs may be detected in restriction maps by noting an alteration of the length of a restriction fragment in which the mutation lies. Base-pair change type of mutations may be detected if their presence inactivates or creates a cleavage site of a particular endonuclease, altering the length of the restriction fragments produced by cleavage in the area of DNA containing the mutation lies.

(C) Restriction Fragment Length Polymorphisms (RPLP)

Different alleles (conditions of a gene) may lead to the production of different proteins and subsequent variation in the phenotype, (the detectable physical,

biochemical, or physiologic makeup of the organism). Variation of DNA within populations is called genetic polymorphism. Even if the polymorphism does not lead to detectable changes in the phenotype by physical appearance or biochemical assays, genetic polymorphism may be detectable by variations in the DNA restriction fragment lengths (RFLP). Polymorphic variation in the restriction map therefore is independent of gene function.

RFLP's have numerous applications including as markers for paternity testing or determining the location of specific genes. For example, mutant genes responsible for inherited diseases such as Huntington's Chorea, a progressive neurological degeneration, have been localized to specific chromosomes in humans by correlating inheritance of RFLP's in families with the inheritance of the particular clinical condition. RFLP patterns of family members who are normal are compared with patterns of family members who are affected with a particular genetic disease.

(D) Other uses for Restriction Endonucleases

Another use of restriction endonucleases is to create and use cloning vectors for the transmission of DNA sequences. For this purpose, the gene of interest needs to be attached to the vector fragment. One way this may be accomplished is by generating complementary DNA sequences on the vector and on the gene of interest so that they can be united (recombined). Some restriction endonucleases make staggered cuts which generate short, complementary, single stranded "sticky ends" of the DNA. An example of such an action is that effected by the EcoR1 endonuclease which cleaves each of the two strands of duplex DNA at a different point.

These cleavage sites lie on either side of a short sequence that is part of the site recognized by the endonuclease. When two different DNA molecules are cleaved with EcoRl the same sticky ends are generated which enables them to combine with each other. The DNA fragment can then be retrieved by cleaving the vector with EcoRl to release the gene.

(E) Exons and Introns

The restriction map of DNA may not correspond directly with the coding sequence of messenger RNA produced by the DNA because DNA sequences of the total gene may consist of exons and introns. Exons are that part of the DNA code that appear in the messenger RNA. Most, but not all, exons code for proteins. Introns are DNA sequences that are usually spliced out of the RNA product before the messenger RNA proceeds to be translated into proteins. Splicing consists of a deletion of the intron from the primary RNA transcript and a joining or fusion of the ends of the remaining RNA on either side of the excised intron. Presence or absence of introns, the composition of introns, and number of introns per gene, may vary among strains of the same species, and among species having the same basic functional gene. Although in most cases, introns are assumed to be nonessential and benign, their categorization is not absolute. For example, an intron of one gene can represent an exon of another. A mosaic gene is defined as one which is expressed through the splicing together of exons carried by one molecule of RNA. In some cases, alternate or different patterns of splicing can generate different proteins from the same single stretch of DNA (Lewin, 1987).

(F) Mitochondrial DNA: Yeast Mitochondria

The DNA contained in mitochondria (cell organelles which contain extranuclear DNA) represents an example of differences arising during evolution between the composition of genes with regard to the exon and intron sequences as well as non-coding sequences. For example, comparing the mitochondrial genes of yeast with those in mammalian systems, indicates that identical mitochondrial proteins are produced despite the disparity in evolution between these species. The yeast mitochondrial genomes, however, are much larger than those occurring in vertebrates due to the absence of introns in the latter and the presence of noncoding spacer DNA in the former.

Primary DNA sequence data are known for many yeast isolates (see deZamaroczy and Bernardi, 1986) in which interstrain differences are due to (i) a small number of large deletions/additions, mainly concerning introns; (ii) a large number of small (10-150 bp) deletions/additions located in the intergenic sequences; (iii) 1-3 bp deletions/additions and point mutations. In Saccharomyces cerevesiae the size of mtDNA can range up to about 84 kilobases, approximately 2/3 of which is non-coding regions. There are more than 20 mitochondria per cell, i.e., approximately 4 genomes per mitochondrion. In comparison, vertebrate mitochondrial DNA is approximately 17 kilobase pairs. In the individual mitochondrion there are usually several copies of a single molecule of DNA. Moreover, there are very many mitochondria per cell. In plants and some unicellular eukaryotes, extranuclear DNA is also found in chloroplasts.

The DNA within the mitochondria directs protein synthesis, just as nuclear DNA does. However, a finite number of proteins are produced. There are general

similarities in the machinery for gene expression in mitochondria of various species, rendering products and information derived in one genus applicable to many others. For some products, the cytochrome c oxidase in yeast, for example, protein synthesis combines factors produced in the cytoplasm encoded by nuclear DNA with those synthesized directly in the mitochondria. Mutations identifying almost all the mitochondrial genes have been detected. Nuclear mutations that interact with, or abolish, the effects of these mutational complexes in the mitochondria, have also been found. Genes coding for many of the same functions are present both in the yeast and the mammalian mitochondrial genome, making the yeast mitochondria a good model system for testing theories on gene expression in higher organisms. The mitochondrial genome of one of the yeast strains, Saccharomyces cerevisiae has provided both information on genetic expression, as well as products which can be useful for analysis of higher systems. (Lewin, 1987: Butow 1985).

In the invention described herein, a new and unique restriction endonuclease has been isolated and purified. One obstacle to purification and characterization of this enzyme in the past has been the inability to accumulate sufficient amounts of the protein, a problem which has been solved by methods disclosed in this invention. A preferred source of the endonuclease described in this invention is yeast mitochondria from a special strain. The endonuclease has wide applications for in vivo or in vitro cleavage of double-stranded DNA from many genera.

In certain aspects, this invention is directed towards a substantially purified endonuclease preparation having an apparent molecular weight of 31,000 daltons as determined by SDS polyacrylamide gel electrophoresis. This endonuclease is capable of cleaving double-stranded

DNA at specific sites (Figure 1). Within these sites, certain specific nucleotides are essential for cleavage, as indicated in Figure 1; at other sites, base substitution is compatible with cleavage.

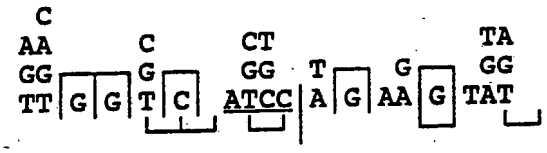
The endonuclease described in this invention is further defined as having a biological activity of up to about 100 units/mg of protein in the crude extract. One unit of endonuclease activity is defined as that amount of enzyme that catalyzes the cleavage of 50ng of a DNA molecule in one hour at 30°C., although other definitions of activity would be within the scope of this invention.

Purification of the endonuclease by standard protein purification techniques was accompanied at each sequential step by increased activity of the endonuclease. Specific examples comprise a biological activity up to about 34,000 units/mg after phosphocellulose chromatography; 50,000 units/mg after Affigel Heparin chromatography; 200,000 units after gel filtration; and 500,000 units/mg after DNA affinity chromatography (Figures 3-6, Table 1).

The endonuclease is translated from a fusion between the upstream exons of the mitochondrial cytochrome oxidase subunit I gene (cox1) of yeast (Figure 1) and the open reading frame (ORF) within the 4th intron (aI4α) of the (cox1) gene. The endonuclease was capable of cleaving recipient DNA molecules near the site of yeast mitochondrial cox1 aI4α intron insertion (Figures 7-8). The endonuclease is also capable of acting as a maturase under certain conditions. Examples of such conditions include presence of a point mutation in the intron reading frame of the mim-2 mutation, (Dujardin et al., 1982) or the presence of the nuclear NAM2 gene (Herbert et al., 1988).

A preferred source of the endonuclease is mitochondria, more specifically yeast mitochondria. An object of this invention was to describe a method of isolating and purifying the endonuclease from yeast mitochondria, more specifically Saccharomyces cerevisiae mitochondria. The WA12 strain of yeast containing mtDNA derived from strain ID41-6/161/PZ27 was a preferred source.

A method is disclosed for preparing the endonuclease. The preferred embodiment comprises culturing yeast that are incapable of splicing the aI4α intron of the cox1 gene, thereby accumulating sufficient amounts of endonuclease to use for isolation, purification, and characterization. The method further comprises preparing mitochondrial extracts from yeast, fractionating the extracts, and selecting a fraction or fractions which contain the endonuclease disclosed in this invention. The endonuclease is further identified by its ability to cleave double stranded DNA at the specific site:



Another object of this invention is a method of cleaving DNA. This method comprises preparing the endonuclease described herein, by the method disclosed in this invention, and incubating the isolated and purified endonuclease protein with DNA so as to effectuate the endonucleolytic cleaving of the DNA.

Other objects and advantages of the invention will become apparent upon reading the following detailed

description and upon reference to the drawings in which aspects of the invention are illustrated as follows:

FIGURE 1: The DNA cleavage site of the aI4α encoded endonuclease.

The cleavage site specificity of the endonuclease is shown. Boxed residues are those believed to be essential for cleavage. The nucleotide changes indicated above the non-boxed nucleotides, are those that have been tested and found to permit cleavage. It is possible that other nucleotides will be identified which can substitute at these and other positions. The enzyme cut sites are indicated by the vertical and horizontal lines.

FIGURE 2: Intron configurations of the *cox1* gene

The *cox1* gene of yeast mtDNA contains up to ten introns, three of which were found recently in cox1 genes from newly studied isolates (Kotylak et al., 1985; Ralph, 1986; Ralph et al., unpublished data). The intron configuration of this gene in strains D273-10B and S. norbensis Y-12,656, is shown in lines 1 and 2, respectively. The *cox1* gene structure of a recombinant resulting from conversion of aI4α is shown in line 3. All three strains shown also contain introns a11 and aI2. Because two of the new introns are present in strains used in this study, a new nomenclature for *cox1* gene introns was used. Previously described introns 3 and 4 are renamed aI3α and aI4α, respectively. One new intron, located between aI3α and aI4α, is called aI3γ (because another new intron, aI3β, is present between aI3α and aI3γ) in S. douglasii [Kotylak et al., 1985]). Exon sequences are indicated by dark boxes whereas introns are designated by open or shaded boxes. Arrows indicate regions sequenced as a basis for this invention. The asterisk denotes the location of two sequence differences

between strains D273-10B and S. norbensis located in exon 5α. The abbreviations used are: B, BamHI; Bc, BCII; H, HaeIII; E, EcoRl (Wenzlau et al., 1989).

FIGURE 3: Purification of the restriction endonuclease by phosphocellulose chromatograrhy

Fraction Ia (see Table I) was dialyzed (M_r cutoff, 10,000) against ice cold buffer containing 50 mM potassium phosphate, pH 7.5, 10% glycerol, 2 mM EDTA, 2 mM DTT (Buffer A) until an ionic equivalent of approximately 100 mM KC1 was achieved. The sample was loaded onto a phosphocellulose column (5 mg protein per packed ml of phosphocellulose), previously equilibrated at a flow rate of 20 ml/hour with Buffer A. The column was washed with 120 ml of Buffer A containing 200 mM KC1 at a rate of 60 ml/hour and then a 240 ml linear gradient was applied using Buffer A containing 200 mM KC1 and 1 M KCI at a rate of 80ml/hour. The aI4α endonuclease activity eluted at 0.7 M KC1. Active fractions were pooled (fraction II, Table I) and dialyzed against Buffer A until an ionic equivalent of 50 mM KC1 was reached.

FIGURE 4: Purification of the restriction endonuclease by affigel herarin chromatography

The dialyzed fraction II (Table I) having endonuclease activity and resulting from phosphocellulose chromatography (Figure 3) was applied to an Affigel Heparin column, previously equilibrated with Buffer A (0.1 mg protein per packed ml of Affigel Heparin), at a rate of 2.5 ml/hour. The column was washed with 15 ml of buffer A containing 50 mM KC1 at 7 ml/hour and eluted with a 60 ml linear gradient of Buffer A containing 100 mM to 600 mM KC1 at a rate of 15 ml/hour. The enzyme eluted at 300 mM KC1, and the active fractions were pooled (fraction III, Table I).

FIGURE 5: Purification of the restriction endonuclease by gel filtration

Fraction III (Table I, Figure 4) was applied to a Sephacryl HS 200 column (1.4 cm x 85 cm) equilibrated with Buffer A containing 100 mM KCI at a rate of 3.2 ml/hour. The column was calibrated with blue dextran 2000, B-amylase, yeast alcohol dehydrogenase, lactate dehydrogenase, boving serum albumin, carbonic anhydrase and cytochrome c. The column was eluted with the Buffer A and 0.95 ml fractions were collected. Fractions containing aI4α endonuclease activity were pooled (fraction IV).

FIGURE 6: Purification of the restriction endonudease by DNA affinity chromatography

Fraction IV (Table I, Figure 5) was applied to a DNA affinity column equilibrated with Buffer A containing 100 mM KCl at 1.5 ml/hour. The column was washed with 6 ml of the same buffer at 6 ml/hour and eluted with 18 ml of a linear gradient of Buffer A containing 100 mM KCl to 800 mM KCl at a rate of 9ml/hour. The aI4α endonuclease activity eluted at 300 mM KCl. The active fractions were pooled and mixed with an equal volume of 50 uM potassium phosphate, pH 7.5, 80% glycerol and 2 mM EDTA and stored at -200C (fraction V, Table I).

FIGURE 7: In vivo double-stranded cuts in recipient DNA molecules

A restriction fragment spanning the intron insertion site was examined for double-stranded breaks in mated cells. The strains 5DSS/D273-10B and COP-19/norbensis were mated, and mtDNA was isolated at the beginning of mating and 3, 6, 9, and 24 hr after mating began. The DNA was digested with either Hpa II (A) or Hae III (B)

and analyzed to detect double-strand cuts near the sites of insertion for omega and al4α, respectively. MtDNA was fractionated on 6% polyacrylamide gels, electrophoretically transferred to Hybond-N membranes, and hybridized to end-labeled oligonucleotides specific to 21S RRNA gene exon sequences (A) or to a sequences (B). A 0.6 kb fragment (arrow) corresponding to an in vivo double-stranded cut near the site of omega insertion was seen from 6-9 hr after initiation of mating in the HpaII digest (A). Similarly, a 0.9 kb fragment (arrow) consistent with a double-stranded cut near the site of al4α insertion was present from 6-9 hr after mating in HaeIII digests (B).

FIGURE 8: Mitochondrial extracts contain an endonuclease that cleaves recipient DNA near the site of intron insertion

Mitochondrial extracts were assayed for their ability to cleave recipient (R) or donor (D) DNA. Plasmids pDR1 (recipient) and pJW1 (donor) were linearized with EcoR1 and 3' end-labeled. These substrates were incubated with mitochondrial extracts from wild-type strain ID41-6/161 and mutant derivatives PZ27, G, and K. Cleavage of pDR1 DNA at or near the site of intron insertion would yield products of 2.1 kb and 3.0 kb (lanes 3, 5, and 9). Cleavage of pJW1 DNA at the equivalent exon site would yield products of 1.2 kb and 2.8 kb. The substrate and extract used are indicated above each lane: 39 ng of extract protein from strain PZ27 was added to lanes 3, 4, and 5; 39 ng and 3.65 μg of extract protein from wild-type strain I41-6/161 was added to lanes 6 and 7, respectively; and 150 ng of extract protein from mutant K and 182 ng of extract protein from mutant G was added to lanes 8 and 9, respectively.

FIGURE 9: Preliminary cleavage site determination of the endonuclease

Single-stranded DNA from phages RS18 and RS19 was labeled using the Sequenase (United States Biochemical Corporation) protocol and the universal sequencing primer. This material was then extended either using dideoxynucleotides (to generate a sequencing ladder) or with unlabeled nucleotides (to generate the substrate). Substrate samples were cleaved with a mitochondrial extract from strain PZ27 (see Figure 7). Figures 9A and 9B show the sequencing gels generated by this protocol using phages RS19 and RS18 as templates, respectively. In each panel, lanes 1-4 represent chain termination reactions with dideoxy G, A, T, and C, respectively; lane 5 is the product of a cleavage reaction on the "end-labeled" substrate; lane 6 shows this region of the gel using uncut substrate to demonstrate no premature primer extension products in the vicinity of the cleavage site. Figure 9C shows the sequence surrounding the cleavage site. Intron sequences are shown in lower case letters while exon sequences are in upper case letters. The large arrow indicates the point of intron insertion. The sites of cleavage on each strand are shown by small arrows. The exon sequences of donor and recipient are nearly identical; two differences in the exon between donor and recipient are shown by asterisks. The bracket sets an upper limit on the boundaries of the recognition site (based on the data of Figure 10).

FIGURE 10: Location of the recognition sequence of the al4α-encoded endonuclease

The recognition sequence was delimited by determining the extent of each sequencing ladder of Figure 9 that was resistant to cleavage by the endonuclease activity. The products of sequencing

reactions using phage RS19 as a template are shown in Figure IOA, lanes 1-4 (terminated with dideoxy G, A, T, and C, respectively). These sequencing ladders were used as a substrate for cleavage reactions using 500 ng of protein extract from strain PZ27 in lanes 5-8. The ladder was resistent to cleavage until a recognition sequence was generated. All chains with further additions were efficiently cleaved. Figure 10B shows the results of a similar experiment using RS18 as a template. Lanes 1-4 show the products of the sequencing ladder (terminated with dideoxy G, A, T, and C, respectively). In lanes 5-8, the products of the ladder were cleaved with extract from strain PZ27. The maximum boundaries of the recognition sequence are shown in Figure 9C.

This invention is directed to a new and unique endonuclease, and to methods of preparing and using this enzyme.

Although other endonucleases have been discovered which cleave DNA at specific sites, there is clearly a need for more endonucleases, in particular those with special and useful properties, to facilitate gene mapping. The endonuclease described herein has the capability of producing relatively larger fragments of DNA and also to act on restriction sites (cleavage sites), that are commonly encountered in the genome. Some enzymes operate only on unusual sequences, e.g., only GC base pairs.

The general objective of this invention is to purify and apply a new endonuclease protein with useful characteristics not previously available to the art. Prior to the invention described herein it had not been possible to isolate and purify the endonuclease protein from its natural source so that its sequence and function could be analyzed.

The endonuclease which is disclosed by the current invention will cleave ordinary DNA sequences. Use of this endonuclease for gene mapping will facilitate and expand the resolution of determination of DNA sequences within fragments cleaved by the endonuclease.

In one of the preferred embodiments, the endonuclease as extracted from mitochondria of strain WA12/PZ27 yields a biological activity of up to about 100 units/mg per protein. These mitochondrial extracts further processed by phosphocellulose chromatography, resulted in the purity increasing to 34,000 units/mg. Application of Affigel® heparin chromatography increased the purity of the endonuclease up to about 50,000 units/mg. Fractionation by gel filtration yielded a biological activity of 200,000 units/mg. Further purification by DNA affinity chromatography yields a biological activity of 500,000 units/mg (Figures 3-6). The purified protein has an estimated molecular weight of 20 kilodaltons as determined by SDS-PAGE.

A maximum of 14 base pair sites are believed to be required for cleavage. There are at least five residues known to be essential for the cleavage; in other positions, some substitutions are permitted and cleavage will still occur (Figure 1).

Most of the endonucleases discovered to date have been from bacterial systems, and most are expressed in E. coli because it is more difficult to purify endonucleases from more complex systems. This limitation was also due to the fact that bacteria are incapable of producing quantities of a specific protein in large enough amounts to provide material on which purification procedures may be employed.

The mitochondrial genomes of many yeast species have been mapped and shown to contain both introns and exons (deZamaroczy and Bernardi, 1986). Two of the most informative loci for genetic investigations are the mosaic genes cob (which codes for cytochrome b) and oxi3 (which codes for subunit I of cytochrome c oxidase. (Figure 2). Many of the introns in these two genes have open reading frames, sequences which are read as extensions of the exon immediately preceding it. An unusual circumstance is that most introns in mitochondrial yeast genomes may be expressed as proteins.

In common with yeast, the mammalian systems have genes for cytochrome b, cytochrome oxidase and subunits of ATPase. The known mitochondrial gene products vary little among eukaryotes as diverse as fungi and vertebrates. The yeast mitochondrial genome consists of large numbers of conserved DNA sequences plus interruptions by an assortment of so called optional sequences. The genome of a given yeast strain appears to be very stable. Coding sequences among strains are conserved, but optional sequences are present or absent at various positions in many combinations. The wide range of functional combinations of introns arose during speciation of yeast. Further combinations of those probably arose in laboratory strains by unselected mitochondrial recombination during the many decades of genetic manipulation in which current laboratory strains were derived (Butow, 1985). The operational definition of optional sequences is that the presence or absence of these sequences do not affect phenotypic expression of the exon coding region.

Much has been learned about the mechanism of RNA splicing by examination of yeast systems. RNA splicing normally removes the introns from the transcripts of genes by breaking the phosphodiester bonds at the

exon-intron boundaries, and subsequently forming bonds between the ends of the free exons. There are various mutants of yeast which are unable to splice the introns; therefore the precursors accumulate in the cells. This can occur in either the nucleus or the mitochondrial genes or both. The mitochondrial introns in yeast may be divided into two types, group 1 and group 2. Group 1 introns appear to have a common secondary structure, short conserved sequences internally, but no specific conservation of sequences at the splice junctions. Some of these introns in the cox1 gene and the single intron of the large mitochondrial ribosomal RNA gene of Saccharomyces cerevisiae, are self-splicing. Mutations in other genes, however, can affect the occurrence of splicing. In both group 1 and group 2 there are some introns which must translate an extensive coding sequence to splice the intron containing it. Some introns encode a protein involved in splicing, called RNA maturases. Maturases appear to result from translation of both exons and introns. Both homologous and non-reciprocal mitochondrial recombinations occur in genetic crosses of yeast. Non-reciprocal exchanges involving some optional sequences have been reviewed by Perlman et al. (1989). The best characterized system of non-reciprocal exchange analyzed in yeast is the omega system. Omega is a 1.1 kb optional intron in the 21S RRNA gene. In genetic crosses between omega positive and omega negative yeast, the progeny are usually omega positive. That is, there seems to be unidirectional transfer of the intron. Mutants lacking this polarity of transfer occur either as base changes close to the site where the intron would be inserted, or as mutations in the intron reading frame preventing its translation. The protein product of the omega intron is not involved in omega splicing.

The hypothesis generated from studying the omega system is that the protein coded by the intron in the

omega positive strain (fit/or omega transposase/ISccI) recognizes the DNA site where the intron should be inserted in an omega negative strain. This causes it to be preferentially inherited. The protein coded by the omega intron has endonuclease activity recognizing a specific sequence in the omega negative gene as a target for cleavage of the doublestranded DNA (Jacquier and Dujon, 1985; Colleaux et al., 1986). The recognition site lies at the site where the intron is inserted. The double-strand break probably initiates what is called a gene conversion, a process in which the sequence of the omega positive gene is copied and replaces the sequence of the omega-negative gene.

Cleavage of intronless (omega-negative) DNA near the site of intron insertion is required for conversion, and it is believed that introns carrying sequences which accomplish gene conversion may have once been independent elements that coded their own splicing or interacted with the DNA of the recipient strain. This creates a transposable element (transposon). Finding that such gene conversions are required for cleavage of the recipient DNA by a site specific endonuclease, leads to the expectation that isolation and purification of such endonucleases would provide extremely valuable not only in understanding the genetic mechanisms of these conversions but also could provide a valuable reagent for gene mapping studies.

When two parental strains having distinguishable mitochondrial DNAs are mated, the progeny are expected to represent the parental DNA in some aspects. Recombination refers to the new union of DNA present in separate parents, in a single progeny DNA. Mitochondrial DNA from different parental strains is capable of recombining.

Through studies of gene conversion in yeast mitochondria, it has been determined that one of the mechanisms achieving gene conversion employs an endonuclease aI4α which cleaves the recipient DNA at specific sites (Wenzlau et al., 1989). For one of these endonucleases, the aI4α-encoded protein, there exist, however, yeast strains that overproduce or overexpress the enzyme relative to wildtype yeast strains because of various mutations of the yeast mitochondrial genome. An example of such a strain of yeast is the WA12 strain containing the PZ27 mutant mitochondrial DNA of Saccharomyces cerevisiae. Strain PZ27 was isolated in 1978. In the mitochondrial genome of this strain the cytochrome oxidase gene (cox1) is present, but the cytochrome b gene is deleted. This configuration blocks splicing of aI4α causing the protein encoded by that intron reading frame to be present at high levels in mitochondria. This protein is a site-specific endonuclease that is translated from a fusion between the upstream exons of cox1 and the open reading frame within the intron aI4α. All yeast have the cox1 gene, and most contain intron aI4α. A few strains, however, have lost the capability of splicing the intron, leading to overproduction of an endonuclease.

In wild-type yeast strains splicing of this intron 4α is assisted by a product of the fourth intron of the cytochrome b gene called the bI4 maturase. Splicing would normally remove the intron. However, there are several strains in which there is an inability to splice aI4α resulting in an accumulation of the endonuclease produced by that region of DNA. This overproduction occurs, for example if there are mutants at the bI4 locus or deletions of the bI4 reading frame. In strain WA12 containing mtDNA from strain PZ27 there is a deletion of the cob gene, consequently a deletion of intron bI4, leading to continued production of the endonuclease

produced by the cox1 gene because there is no splicing of the intron. Otherwise the strain has a wild-type cox1 gene based on complementation studies.

Some other strains exist with defects in the bI4 maturase that would lead to this overproduction. In fact any strain lacking bI4 maturase and thereby unable to splice out aI4α will lead to overproduction of the endonuclease. A preferred embodiment is to use the WA12 strain as a source of crude extracts of mitochondria. After culturing the yeast, mitochondrial extracts can be prepared by standard methods of lysing cells and fractionation. These extracts can be further purified by being subjected to phosphocellulose chromatography, affigel heparin chromatography, gel filtration and DNA affinity chromatography. Any other suitable methods of purification of proteins known to those skilled in the art may be applied.

The endonuclease described herein was capable of cleaving double-stranded DNA at a specific site to achieve gene conversion in recipient strains. (Figures 7-8). It also has the potential to become a maturase under certain conditions. A maturase is an intron-encoded RNA splicing protein. It has been reported that the maturase coded in yeast by the fourth intron of the cytochrome b gene (bI4) is essential for splicing the fourth intron of the gene encoding subunit 1 of the cytochrome c oxidase (cox1 gene). The cox1 intron 4α encodes a protein that is closely related structurally to the bI4 maturase, but in wild-type yeast strains is not utilized in splicing either the fourth intron of cox1 or cob. However, the aI4α product can be activated to a new maturase form by either a point mutation (mim-2) (Dujardin et al., 1982) or the presence of particular mutant forms of the nuclear NAM2 gene (Herbert et al., 1988).

Although intron-encoded endonucleases have not previously been isolated and purified from yeast mitochondria, the use of special laboratory strains in this invention which continue to produce these products due to lack of splicing, permitted the isolation of endonucleases in amounts sufficient to be isolated, purified and characterized. The function and structure of one of these endonucleases was determined for the first time as an aspect of this invention because of the increased amounts obtained from the special strain WA12/PZ27 of Saccharomyces cerevisiae. The specific site of cleavage of double-stranded DNA was determined by presenting known sequences of DNA of different lengths to the endonuclease and studying the resulting cleavage products. A partial recognition site determination is shown in Figure 10.

The purified fraction is then selected which comprises an endonuclease capable of cleaving DNA at specific sites. This capability can be tested by various assays. The cleavage site, plus flanking DNA sequences can be introduced into various plasmids to use as an in vitro assay. Essentially any plasmid DNA containing the cleavage site will be suitable. In a preferred embodiment the assay used comprised a plasmid pRSX which contains a 292 base pair, RsaI-HindIII fragment that includes the cleavage site from Saccharomyces capensis cloned into the vector PBS(+) from Stratagene.

Two assays were used to determine the activity of the endonuclease:

1. A supercoiled DNA preparation of pRSX was incubated with the endonuclease, and cleavage activity was determined by gel electrophoresis of the subsequent fragments. If cleavage has occurred there will be conversion of the supercoiled DNA to the linear form

of the plasmid which can be determined by gel electrophoresis;

2. pRSX DNA was linearized with ScaI. Using standard techniques the linear plasmid was end-labelled, extracted, and endonuclease activity was determined by the production of two radioactive fragments after incubation with the enzyme for which activity is to be determined.

An object of this invention is to cleave DNA at specific sites. Determining whether this was achieved by preparing the endonuclease protein as defined above, was done by incubating the enzyme with DNA so as to effectuate the cleaving of DNA.

Strains shown in Figure 2 were used in crosses to study the transmission and recombination behavior of three group 1 introns of the cox1 gene. The data shown in Figure 7 indicated that the mated cells contain an endonuclease activity capable of cleaving recipient genomes in the vicinity of the site of conversion. Experiments employing petite mutants of yeast which do not express mitochondrial genes, showed that no conversion occured in petites. The conclusion from these studies was that a product of mitochondrial protein synthesis is needed for aI4α conversion. Results of crossing strains of yeast differing in their intron configuration of three mosaic genes (cox1, cob, 21S rRNA), and use of various strains having point mutations of aI4α, led to the conclusion that the aI4α intron likely encodes the information needed for aI4α conversion. The aI4α reading frame must be intact to effect conversion.

In this invention, purification and subsequent application of an enzyme not previously purified, was

achieved and found to cleave DNA as shown in Figure 1. The endonuclease requires Mg^{2+} at a concentration of about 25mM for activity and shows high activity at pH 7.5 in the presence of 200mM salts of monovalent cations. Activity does not appear to be stimulated by the addition of nucleoside triphosphates (see also Figure 8). Other properties of the endonuclease illustrated in Figure 8 are: lane 3: the mitochondrial extract from strain PZ27 cleaves the recipient plasmid once near the intron insertion site (this linear product can be circularized by DNA ligase); lane 4: donor DNA is not cleaved; lane 5: only recipient DNA is cleaved; lanes 6 & 7 show that cleavage activity is not detected in extracts of mitochondria from a wild-type strain capable of efficient splicing of aI4α; lane 8: no cleavage using extract from an aI4α mutant with a truncated intron ORF; lane 9: cleavage from an extract of an aI4α mutant with an intact intron ORF.

The endonuclease has been shown to be active in E. coli by use of an artificially engineered sequence of the endonuclease coding sequences in a transforming vector, using the universal genetic code. However, it is highly toxic in E. coli probably due to its cleavage of double stranded E. coli DNA (Delahodde et al., 1989).

A major advantage of the purified endonuclease, is that it cuts DNA relatively infrequently, and consequently is particularly useful for genome mapping. If all 14 nucleotides are required for cleavage, then one can estimate 1 to 2 cleavage sites per human chromosome. However, because there is some relaxation in the specificity of the cleavage site, empirical tests leading to actual determination of cuts effected by the endonuclease in clonal portions of human chromosome number 3, show one site about every 200 kilobase pairs.

Since the DNA sequence which the inventors have discovered to encode for the endonuclease is available (Bonitz et al. 1980), it could be engineered for expression in, e.g., E. coli (Delahodde et al. 1989). However, the preferred embodiment is to prepare this endonuclease from yeast mitochondrial extracts because of possible variations in the protein during the course of molecular engineering and expression in other species and because of its apparent toxicity to bacterial cells.

A major current scientific project is to map the entire human genome. This is an expensive, labor-intensive, and difficult task which will be facilitated by use of the endonuclease described in this invention. Human gene mapping permits detection of carriers of abnormal genes, and isolation and cloning of genes which cause abnormal human conditions. A goal of the mapping project is further elucidation of gene action, paving the way for therapy. Complete mapping of the human genome will also contribute to the field of preventive medicine by making it possible to determine the susceptibility of individuals to specific genetic disorders. Because the phenotype (expression of such genes) represents the interaction between genes and environment, preventive action based an knowledge of the genetic complement of the individual may be achieved by altering their environment (diet, drug, medication) to reduce or eliminate the severity of the abnormal genetic defects. The ultimate goal, of course, is to identify these genes so that direct gene therapy will be possible. If the genomic sequences are known, then normal genes may be directed into the genotype, perhaps at early stages of development, to counteract the effects of abnormal genes. A further goal of gene mapping is to understand the interaction and interrelationships of genes on different parts of the chromosome and on different chromosomes.

The purification, isolation and use of the endonucleasd which is an object of this invention are further described in the following examples:

Example 1: Yeast Strain. Purification of the endonuclease is made possible by the use of a strain or strains that overproduce the enzyme. In wild-type yeast strains, splicing of aI4α is assisted by the "maturase" product of the fourth intron (bI4) of the cytochrome b gene (cob); mutants of the bI4 maturase, or deletions of all or portions of the bI4 reading frame, result in a block in splicing of aI4α. Because the mRNA for aI4α endonuclease is a fusion of the intron reading frame with those of the upstream exons, the inability to splice aI4α results in an accumulation of the endonuclease mRNA and, hence, an overproduction of the protein. Therefore, to facilitate purification of the endonuclease protein, a strain (WA12), containing the mutant mitochondrial genome, PZ27, which is deleted for the cob gene, is used. (Note, however, that any strain with a defect in the synthesis of the bI4-encoded maturase would overproduce the aI4α endonuclease and therefore be suitable as a starting strain for purification.)

Example 2: DNA Cleavage site. The cleavage site for the endonuclease is the sequence shown in Figure 1. This site is cleaved with high efficiency by crude and purified preparations of the endonuclease which catalyzes a 4 bp staggered double strand break as shown in Figure 1, leaving 3' OH overhangs. Preliminary determination of the cleavage site is shown in Figure 9. The boundaries of the recognition site are shown in Figure 10. Methods are detailed in Wenzlau et al., 1989.

Example 3: Assays for endonuclease activity. For an in vitro assay of endonuclease activity, the cleavage site shown in Figure 1, plus flanking DNA sequences was

introduced into various plasmids. Essentially any plasmid DNA containing the cut site is suitable. For the assays described herein, the plasmid pRSX containing a 292 bp Rsa I - Hind III fragment that includes the cut site from S. capensis cloned into the vector pBS (+) from Stratagene was used.

Two assays were conveniently used to follow enzyme activity:

a) Cleavage of supercoiled DNA. A supercoiled preparation of pRSX was incubated with the endonuclease under the assay conditions described below, and cleavage activity was evident, following gel electrophoresis of the sample, as a conversion of the supercoiled to the linear form of the plasmid;

b) Linearization and end-labelling of pRSX. pRSX DNA was linearized with ScaI. After phenol extraction and ethanol precipitation, the linear plasmid was end-labelled with ^{32}P dATP and T4 DNA polymerase by standard procedures. The labelled DNA was then extracted by phenol and the unincorporated dNTPs were removed by gel filtration. The radioactive pRSX DNA was ethanol precipitated and resuspended in sterile water. Endonuclease activity was followed by the production of two radioactive fragments following incubation with active enzyme.

The endonuclease assay conditions used were as follows: Reaction mixtures (0.025 ml) contained 25 mM Tris-HCI, pH 7.5, 20 mM $MgCl_2$, 100 mM KC1, 2 mM DTT and 100 ng pRSX plasmid DNA. Incubation was carried out at 30°C for 30 minutes. Reactions were stopped by being placed on ice and a 1/10 volume gel loading buffer was added. The samples were analyzed by electrophoresis in 0.7% agarose gels containing 1 μg/ml ethidium bromide.

For purposes of this invention, one unit of endonuclease activity was defined as that amount that catalyzes the cleavage of 50 ng of the DNA substrate in one hour at 30°C. Other definitions used in the art would be consistent with this invention.

Example 4: Protocol for the Purification of the Endonuclease. The following table, Table I, is an overview of the protocol which was employed to purify the enzyme, and also indicates the degree of purification that was achieved at each step.

Table I: SUMMARY OF PURIFICATION OF THE RESTRICTION ENDONUCLEASS

	Sequential Purification Steps	Activity* U X 10^{-3}	Yield	Protein (mg)	Specific Activity (U/mg)
I.	Crude Extract	26	100	216	100
Ia.	40-60% $(NH_4)_2SO_4$	3	--	114	--
II.	Phosphocellulose	20	79	0.6	34,000
III.	Affigel Heparin	5	19	0.1	50,000
IV.	Gel Filtration	4	16	0.02	200,000
V.	DNA Affinity	1	4	0.002	500,000

*For purposes of this invention, one unit of endonuclease activity was defined as that amount that catalyzed the cleavage of 50 nanograms of DNA substrate in one hour at 30°C. Other definitions of activity are within the scope of this invention.

(a) Preparation of Crude Mitochondrial Extract and Ammonium Sulfate Fractionation

A Saccharomyces cerevisiae strain (WA12/PZ27) was grown in medium composed of 1% yeast extract, 1% bacto-peptone and 2% galactose to late log phase at 30°C. The cells were collected by centrifugation, washed with 2

mM EDTA, at pH 8, and resuspended at 2 ml/g, wet weight, in buffer containing 33 mM potassium phosphate, pH 5.8, 1 M sorbitol and 2mM EDTA. Mureinase was added at a ratio of 4 mg/g, wet weight, and the mixture was incubated on a platform shaker at room temperature for one hour. Spheroplasts were collected by centriguation at 1000 x g for 10 minutes at 3°C and resuspended in ice cold breakage buffer which contained 50mM Tris-HCl, pH 7.5, 0.6 M sorbitol, 2 mM EDTA, 2mM DTT and protease inhibitors (which were also included in all subsequent steps of the purification): PMSF (1 mM), leupeptin (1 μg/ml) and aprotinin (1 μg/ml). Spheroplasts were lysed by shaking with glass beads. The lysate was recovered and the glass beads were washed 3 times with breakage buffer. The combined spheroplast lysate was centrifuged at 1000 x g for 10 minutes at 3°C.

A high salt extract of the crude supernatant was prepared by addition of 0.33 volume of 4 M KCl to achieve a final concentration of 1 M KCl. $MgCl_2$ and DTT were added to final concentrations of 10 mM and 2 mM, respectively. The high salt extract was stirred in the cold for 30 minutes and centrifuged at 15,000 x g for 15 minutes at 3°C. The supernatant (fraction I, Table I) was recovered and diluted 2-fold by addition of ice cold buffer containing 50 mM Tris-HCl, pH 7.5, 2 mM EDTA and 2 mM DTT. One-tenth volume of a 25% streptomycin sulfate solution was added and the extract was stirred in the cold for 30 minutes followed by centrifugation at 15,000 x g for 15 minutes at 3°C.

Solid ammonium sulfate was added to the crude supernatant over a period of 30 minutes to achieve 40% of saturation at 0°C. Stirring was continued for another 30 minutes to achieve 60% of saturation at 0°C. Stirring was continued for another 30 minutes and the precipitate was collected by centrifugation and discarded. The

supernatant was recovered and additional solid ammonium sulfate was added over a period of 30 minutes and the precipitate, which contained the aI4α endonuclease activity, was collected by centrifugation at 15,000 x g for 15 minutes at 3°C and resuspended in buffer containing 10 mM potassium phosphate, pH 7.5, 45% glycerol and 2 mM EDTA and stored at -20°C (fraction Ia in Table I).

The next step in the purification was to subject fraction Ia (Table I) to phosphocellulose chromatography. The details of the procedure are described in Figure 3. Fractions showing endonuclease activity were pooled (fraction II in Table I) and dialyzed against Buffer A until an ionic equivalent of 50mM KCl was reached. This dialyzed fraction was then applied to an Affigel Heparin column (Figure 4). The active fractions from this procedure were pooled to form Fraction III (Table I). This pooled fraction was then applied to a Sephacryl HS200 column (Figure 5). The active fractions obtained from gel filtration were pooled to form Fraction IV (Table I). This fraction was applied to a DNA affinity column (Figure 6). The pooled active fractions (V, Table I) were mixed with an equal volume of 50mM potassium phosphate, pH7.5, 80% glycerol and 3mM EDTA and stored at -20°C.

Example 5: Analysis of mtDNA. Mitochondrial DNA was prepared by the procedure of Bingham and Nagley (1983). Restriction enzyme digestions were performed as recommended by the enzyme supplier (Bethesda Research Laboratories). Fragments were separated on 1% agarose or 6% polyacrylamide gels and transferred by the well-known method of Southern (1975) or electrophoretically to Hybond-N membranes (Amersham). Hybond-N membranes were prehybridized in 6x SSC containing 10x Denhardt's, 0.5% SDS, and 0.5 mg/ml of calf thymus DNA at 65°C.

Oligonucleotides were end-labeled with T4 polynucleotide kinase and allowed to hybridize to membranes for 2 hr at 42°C. Blots were washed three times for 15 min each at 50°C in 6x SSC. Other probes, generated by the multiprime DNA labeling system (Amersham), were hybridized to membranes overnight at 42°C. Blots were washed twice at 42°C for 15 min each in 2x SSC, containing 0.1% SDS, and twice for 15 min each in 0.2x SSC, containing 0.1% SDS. All filters were exposed to Kodak XAR-5 X-ray film.

Example 6: Extract Preparation and Endonuclease Assay. Galactose-grown cells were harvested at late logarithmic phase, and a mitochondrial fraction was isolated as described by Hudspeth et al. (1980). All further manipulations were done at 4°C. Mitochondria were resuspended in a solution containing 0.6 M sorbitol, 25 mM Tris-HCl (pH 7.5), 25 mM EDTA, and 30% Percoll and were further purified by banding in self-forming Percoll gradients obtained by centrifuging the resuspended mitochondria at 15,000 x g for 20 min. The mitochondrial band was isolated and diluted with 10 vol of a solution containing 0.5 M NH_4Cl and 25 mM Tris-HCl (pH 7.5) and centrifuged at 15,000 x g to pellet the mitochondria. Mitochondria were lysed in a solution containing 50 mM Tris (pH 7.5), 100 mM NH_4Cl, 10 MM $MgCl_2$, 2 mM DTT, and 1% NP40. The extract was clarified at 15,000 x g, and the supernatant was used for endonuclease assays. Protein concentration was determined by the BCA protein assay (Pierce Chemical Company).

For the cleavage experiment described in Figure 8, the substrates pDR1 and pJW1 were linearized with EcoR1 and the overhangs filled in with [α-^{32}P]dATP and dTTP using the Klenow fragment. Unincorporated nucleotides were eliminated on a spin column and the DNA was phenol extracted, ethanol precipitated, and resuspended in a

solution containing 10 mM Tris (pH 7.5) and 1 mM EDTA. Reactions were performed with 10 ng of substrate in a 25 µl reaction. Reaction conditions for this study were 25 mM Tris(pH 7.5), 25 mM $MgCl_2$, 200 mM NH_4Cl, and 2 mM DTT at 30°C. Reactions were allowed to proceed for 5 min and were stopped with 2.5 µl of a solution containing 50 mM Tris (pH 7.5), 100 mM EDTA, 2% SDS, 15% Ficoll, and 1% bromophenol blue. The reaction products were separated on a 1% agarose gel and visualized by autoradiography.

Example 7: Mapping of the Cleavage and Recognition Sites. Using a modification of the Sequenase® (United States Biochemical Corporation) sequencing strategy, the cleavage site was mapped by generating double-stranded substrates from single-stranded templates (phages RS18 and RS19) bearing the cleavage site. Two nanograms of the universal sequencing primer was annealed to 1 µg of each template in a 20 µl Sequenase® labeling reaction. The reaction was extended with 1 mM dNTPs to generate the "end-labeled" substrate. One-fifth of that material was cleaved with PZ27 extract as described above. An aliquot (1/50) was separated alongside a sequencing ladder generated from the rest of the labeling mixture and separated on 5% polyacrylamide gels. To determine the boundaries of the recognition site, cleavage reactions were performed on the sequencing ladders themselves (Figures 9 and 10).

. . .

While the invention is susceptible to various modifications and alternative forms, specific embodiments thereof have been shown by way of example in the drawings and have been described in detail. It should be understood, however, that it is not intended to limit the invention to the particular forms disclosed, but on the contrary, the intention is to cover all modifications,

equivalents, and alternatives falling within the spirit and scope of the invention as defined by the appended claims.

REFERENCES

1. Bingham, C.G. and Nagley, P. A petite mitochondrial DNA segment arising in exceptionally high frequency in a mit - mutant of Saccharomyces cerevisiae, Biochem. Biophys. Acta. 740:88 (1983).

2. Bonitz, S. E., et al. Assembly of the mitochondrial membrane system. Structure and nucleotide sequence of the gene coding for submit I of yeast cytochrome oxidase, J. Biol. Chem. 255:11927-11941 (1981).

3. Butow, R.A. Non-reciprocal exchanges in the yeast mitochondrial genome, Trends Genet. 1:81 (1985).

4. Butow, R.A., et al. The unusual var1 gene of yeast mitochondrial DNA, Science 228: 1496 (1985).

5. Colleaux, L., et al. Universal code equivalent of a yeast mitochondrial intron reading frame is expressed into E. coli as a specific double strand endonuclease, Cell 44:521 (1986).

6. Delahodde, A., et al. Site specific DNA endonuclease and RNA maturase activities of two homologous intron-encoded proteins from yeast mitochondria, Cell 56:431 (1989).

7. Dujardin, G., et al. Single base substitution in an intron of oxidase gene compensates splicing defects of the cytochrome b gene, Nature 298:628 (1982).

8. Hanson, D.K., et al. Evidence for translated intervening sequences in the mitochondrial genome of Saccharomyces cerevisiae, J. Biol. Chem. 257:3218 (1982).

9. Herbert, C.J., et al., The NAM2 protein from Saccharomyces cerevisiae and S. douglassi, are mitochondrial leucyl tRNA synthetases and are involved in mRNA splicing, EMBO J 7:473 (1988).

10. Hudspeth, M., et al. Rapid purification of yeast mitochondrial DNA in high yield. Biochem. Biophys. Acta. 610:221 (1980).

11. Jacquier, A., Dujon, B. An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene, Cell 41:383 (1985).

12. Kotylak, Z., et al. Intron encoded proteins in mitochondria: key elements of gene expression and genomic evolution. In Achievements and Perspectives

of Mitochondrial Research, v. II, Biogenesis, Quagliaiello, E.C. et al. (eds.), Elsevier science Publishers B.V., Amsterdam, p.1 (1985).

13. Lewin, B. Genes III Chap. 4, Chap. 18, Wiley and Sons, N.Y. (1987).

14. Perlman, P.S., et al. Mosaic genes in yeast mitochondria. In Gene Structure and Expression, Dean, D.B. et al.(eds.), The Ohio State University Press, Columbus, Ohio, p. 1 (1980).

15. Perlman, P.S. and Butow, R.A., Science, 246:1106 (1989).

16. Ralph, D. Evolution of cytoplasmic genomes. Ph. D. Thesis. The Ohio State University, Columbus, Ohio (1986).

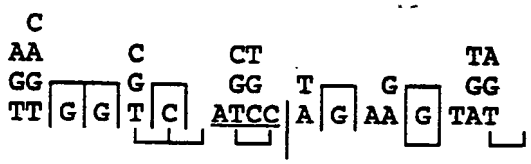
17. Wenzlau, J.M., et al. A latent intron-encoded maturase is also an endonuclease needed for intron mobility, Cell 56:421 (1989).

18. deZamaroczy, M., Bernardi, G. The primary structure of the mitochondrial genome of Saccharomyces cerevisiae - A review, Gene 47:155 (1986).

19. Zassenhaus, H.P. Genetic and biochemical characterization of mitochondrial mutants in the var1 region of Saccharomyces cerevisiae, Ph.D. Thesis, The Ohio State University (1979).

CLAIMS:

1. An endonuclease protein having an apparent molecular weight of about 29,000 daltons as determined by SDS-PAGE, and which is capable of cleaving double-stranded DNA at a specific site identified as follows:



2. The endonuclease protein of claim 1, further defined as a mitochondrial-derived enzyme.

3. The endonuclease protein of claim 2, further defined as a mitochondrial enzyme derived from yeast.

4. The endonuclease protein of claim 3, further defined as a Saccharomyces cerevisiae-derived enzyme.

5. The endonuclease of claim 1 further defined as having a biological activity of up to about 100 units/mg protein in crude extracts.

6. The endonuclease of claim 5 further defined as having a biological activity of up to about 34,000 units/mg after phosphocellulose chromatography.

7. The endonuclease of claim 6 further defined as having a biological activity of up to about 50,000 units/mg after affigel heparin chromatography.

8. The endonuclease of claim 7 further defined as having a biological activity of up to about 200,000 units/mg after gel filtration.

9. The endonuclease of claim 8 further defined as having a biological activity of up to about 500,000 units/mg after DNA affinity chromatography.

10. The endonuclease of claim 1 wherein the endonuclease is translated from a fusion between the upstream exons of the mitochondrial cytochrome oxidase subunit I gene (cox1) of yeast and the open reading frame (ORF) within the 4th intron (aI4α) of the cox1 gene.

11. The endonuclease of claim I wherein the enzyme is capable of cleaving recipient DNA molecules near the site of yeast mitochondrial cox1 intron (aI4α) insertion.

12. The endonuclease of claim 1 which is also capable of acting as a maturase under certain conditions, said conditions comprising the coincidence of a point mutation in the intron reading frame (the mim-2 mutation), or the presence of the nuclear NAM2 gene.

13. A method for preparing an endonuclease having the capability of cleaving double stranded DNA, comprising the steps:

culturing yeast that are incapable of splicing the aI4α intron of the cox1 gene;

preparing a mitochondrial extract from the yeast;

fractionating the extract; and

selecting a fraction or fractions which comprises an endonuclease as defined by claim 1.

14. The method in claim 13 wherein the yeast comprises the WA12/PZ27 strain of Saccharomyces cerevisiae.

15. The method in claim 13 wherein the yeast comprises a strain incapable of expressing the bI4 maturase activity of the cytochrome b gene (cob) in an appropriate culture media.

16. A method of cleaving DNA comprising the steps of:

preparing an enzyme as defined by claim 1; and

incubating the enzyme with DNA so as to effectuate the endonucleolytic cleaving of the DNA.

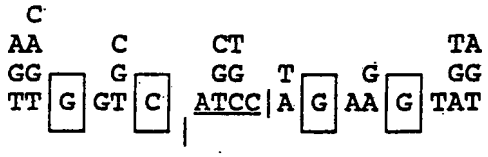


FIGURE 1

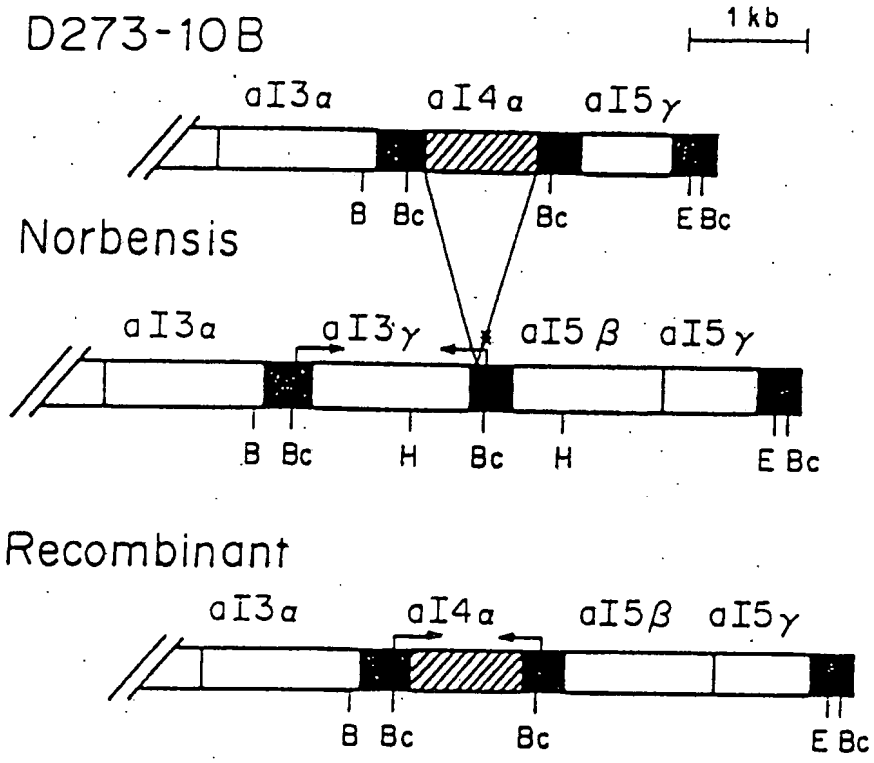

D273-10B
1 kb
αI3α
αI4α
αI5γ
B Bc
Bc
E Bc
Norbensis
αI3α
αI3γ
αI5β
αI5γ
B Bc
H
Bc
H
E Bc
Recombinant
αI3α
αI4α
αI5β
αI5γ
B Bc
Bc
E Bc

FIGURE 2

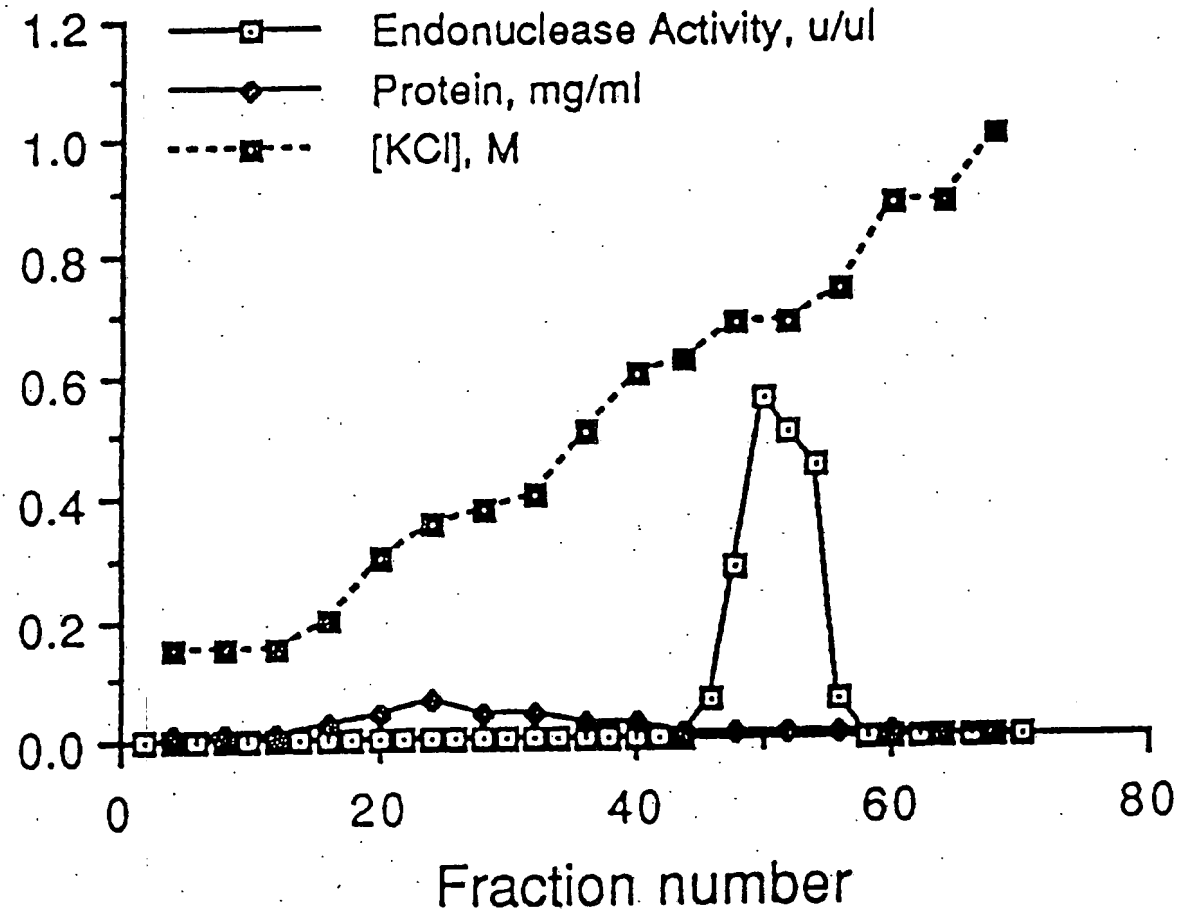

Endonuclease Activity, u/ul
Protein, mg/ml
[KCl], M
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
Fraction number

FIGURE 3

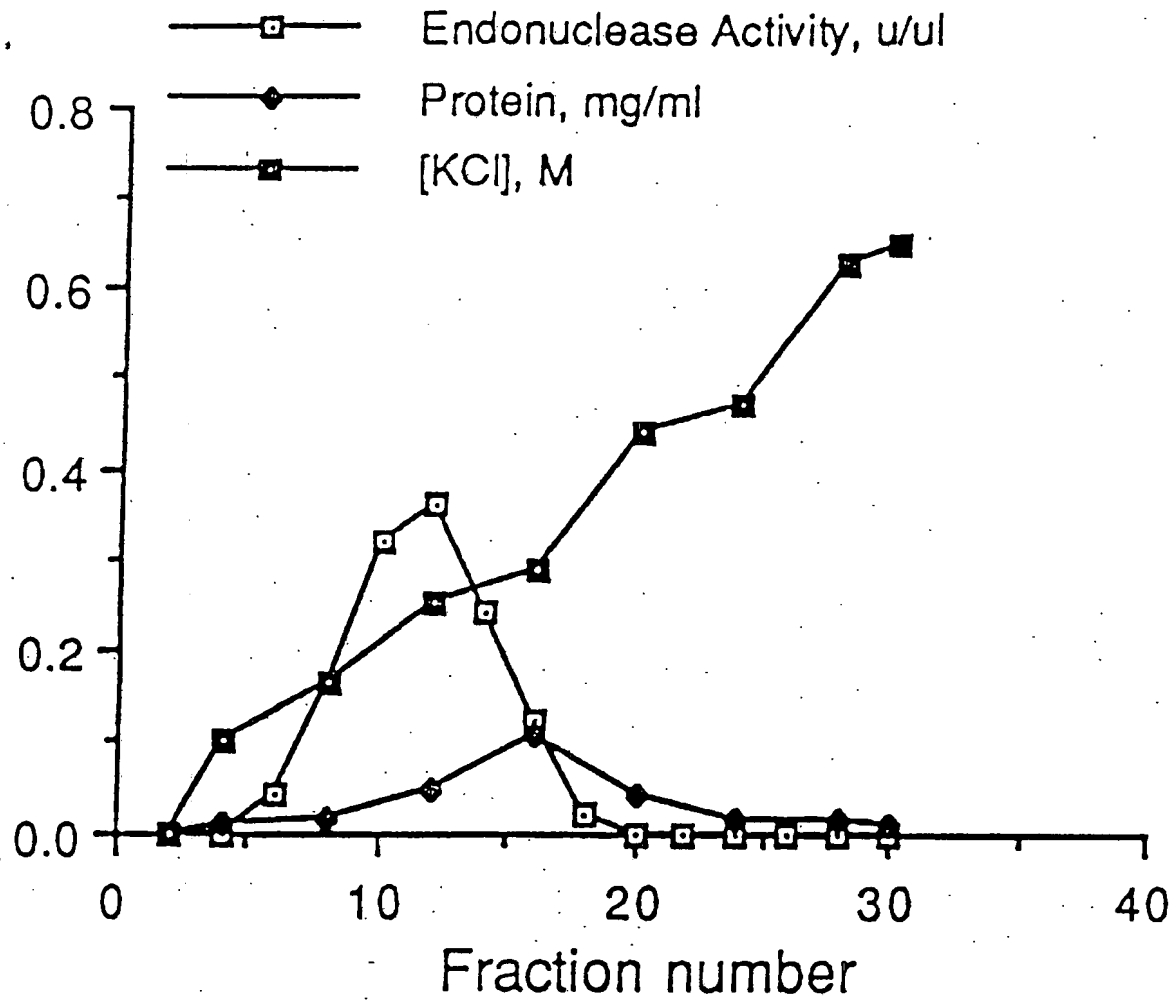

Endonuclease Activity, u/ul
Protein, mg/ml
[KCl], M
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
Fraction number

FIGURE 4

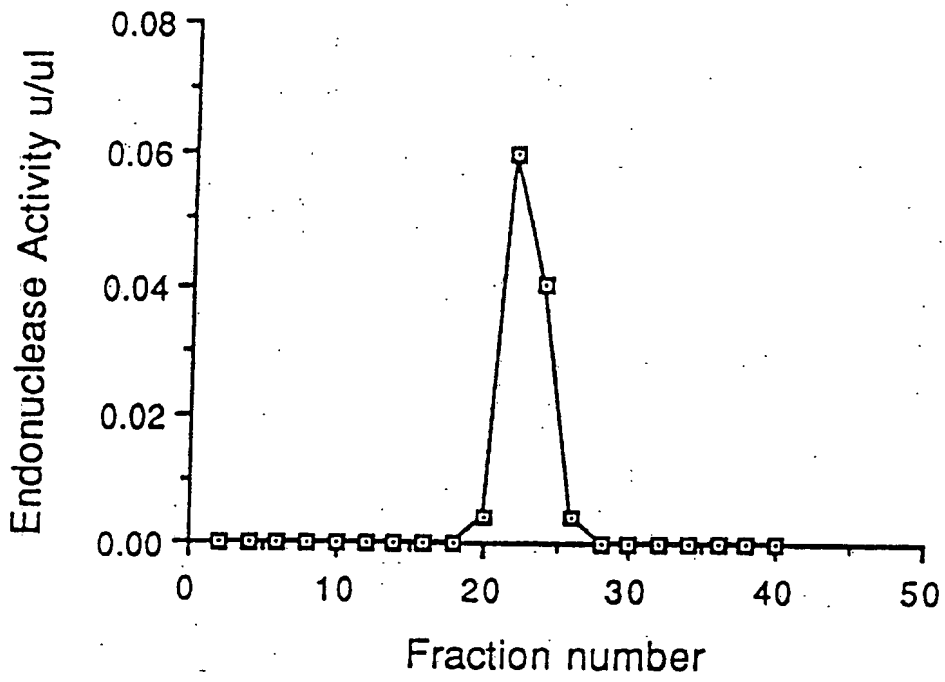

Endonuclease Activity u/ul
0.08
0.06
0.04
0.02
0.00
0
10
20
30
40
50
Fraction number

FIGURE 5

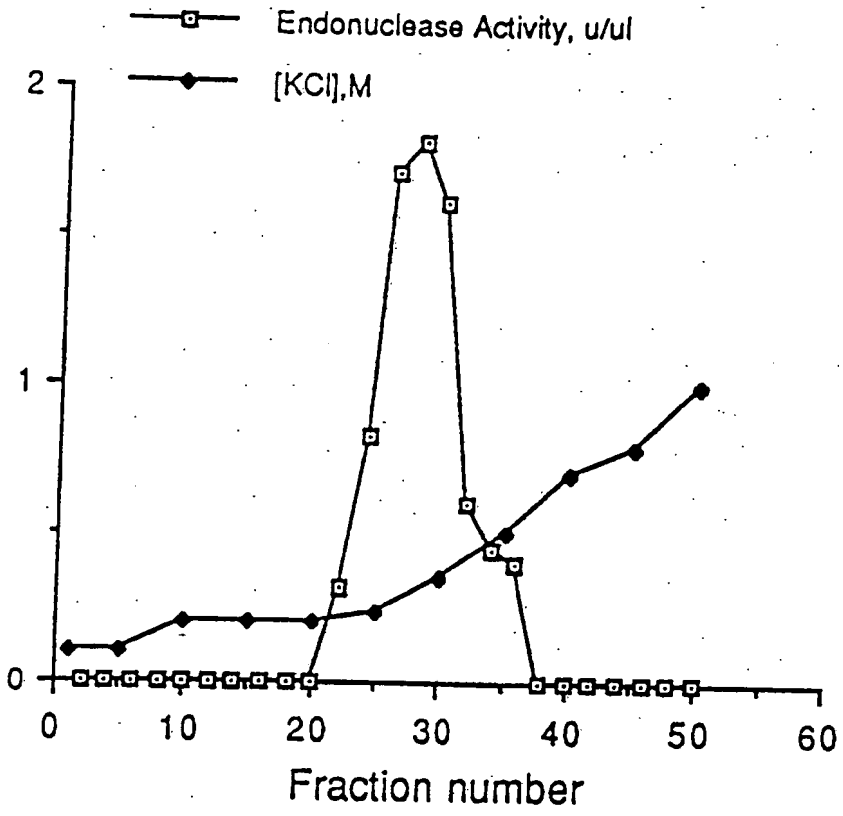

Endonuclease Activity, u/ul
[KCl],M
2
1
0
0
10
20
30
40
50
60
Fraction number

FIGURE 6

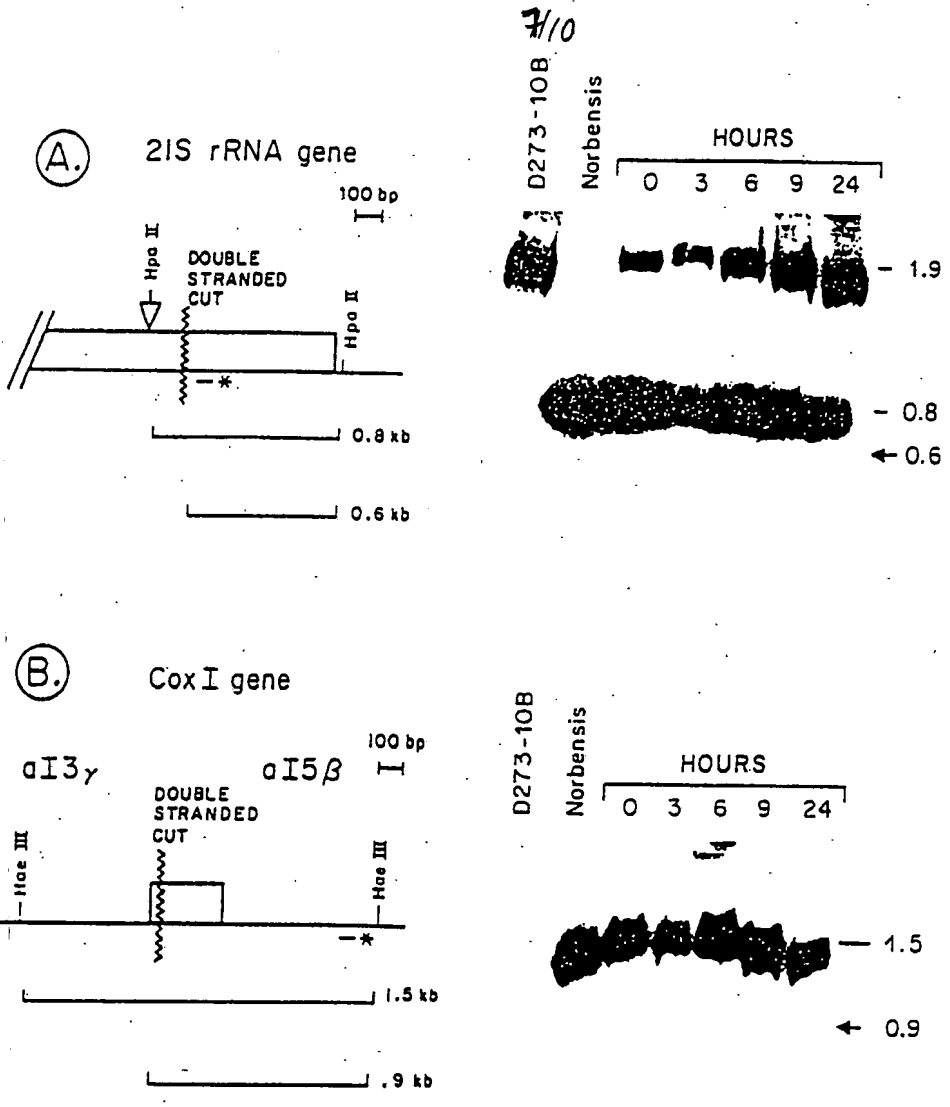

7/10
A.
21S rRNA gene
100 bp
Hpa II
DOUBLE STRANDED CUT
Hpa II
0.8 kb
0.6 kb
D273-10B
Norbensis
HOURS
0 3 6 9 24
1.9
0.8
0.6
B.
CoxI gene
aI3γ
aI5β
100 bp
DOUBLE STRANDED CUT
Hae III
Hae III
1.5 kb
.9 kb
D273-10B
Norbensis
HOURS
0 3 6 9 24
1.5
0.9

FIGURE 7

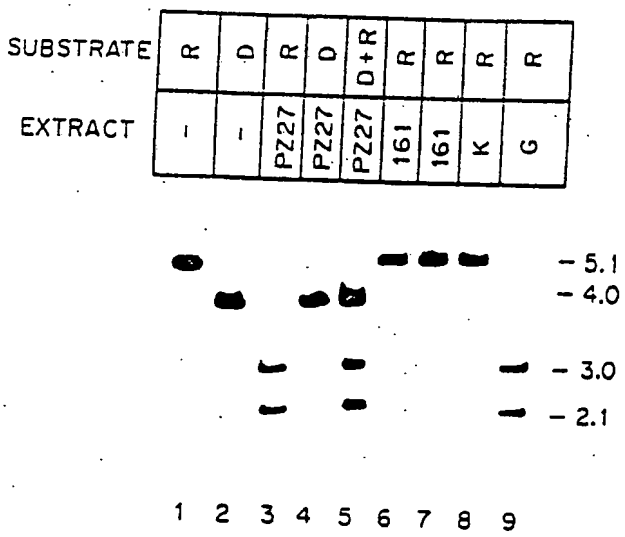

SUBSTRATE
EXTRACT
R
D
R
D
D+R
R
R
R
R
-
-
PZ27
PZ27
PZ27
161
161
K
G
- 5.1
- 4.0
- 3.0
- 2.1
1 2 3 4 5 6 7 8 9

FIGURE 8

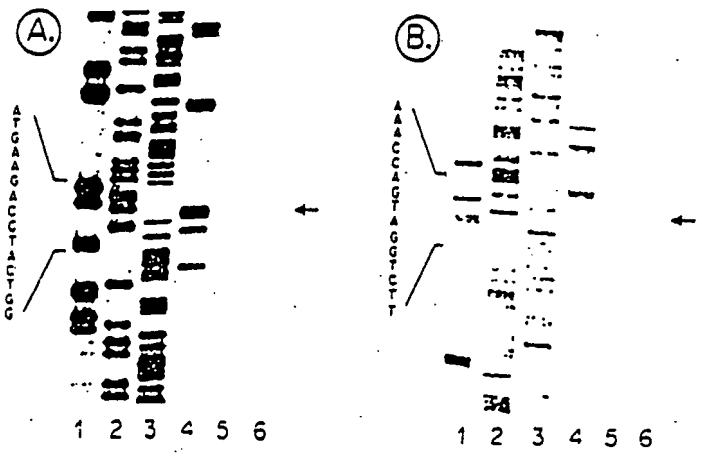


FIGURE 9

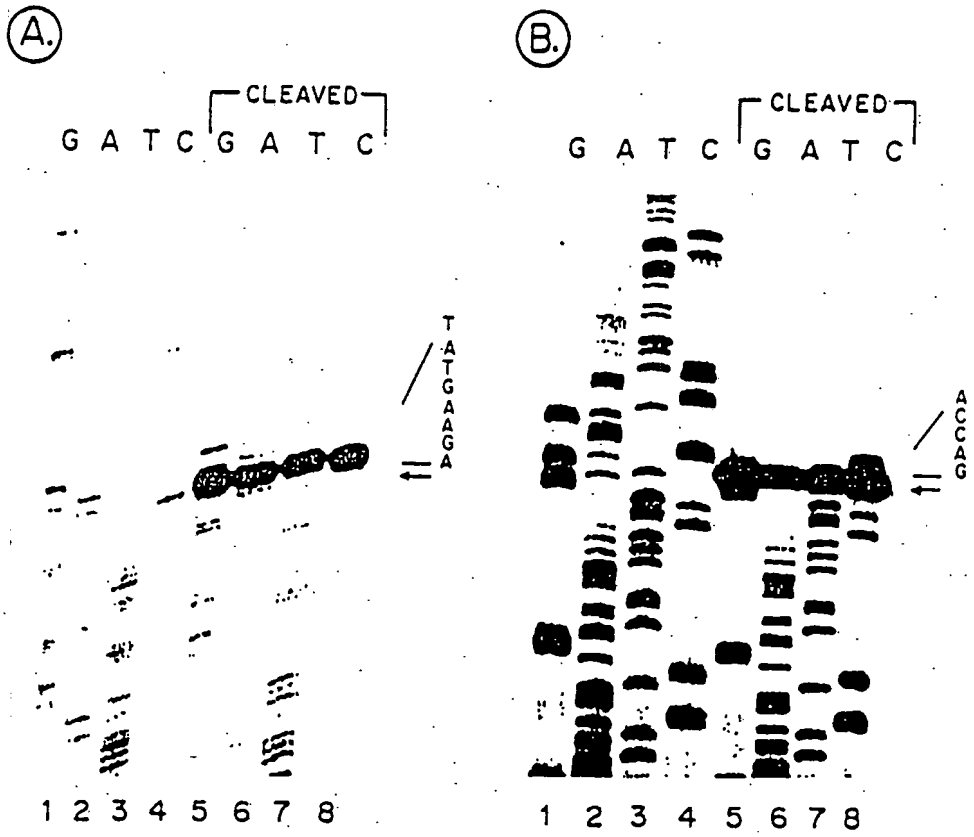

A.
CLEAVED
G A T C G A T C
TATGAAGA
1 2 3 4 5 6 7 8
B.
CLEAVED
G A T C G A T C
ACCAG
1 2 3 4 5 6 7 8

FIGURE 10

Example 4

Quantification of Normal KIT and Splice Mutant KIT (Intron 17 nt1 $^{G\rightarrow A}$)

As the splice site mutation is present in only one of the duplicated regions of *I* and not in the duplicated region of I^P, the various genotypes can be expected to have the attributes described in Table 3.

TABLE 3

Genotype	Copies of Normal KIT	Copies of KIT containing the splice mutation	Ratio of normal KIT to splice mutant KIT
I/I	2	2	1:1
I/i	2	1	2:1
i/i	2	0	2:0
I/I^P	3	1	3:1
I^P /i	3	0	3:0

Due to the dominance of allele *I*, three of the genotypes in Table 2 are carried by white animals and therefore can not be identified by phenotypic characterisation. Quantification of the relative amounts of the normal *KIT* gene and the splice mutant *KIT* gene allows the ratio between the two to be calculated, and therefore the genotype of individual animals predicted. This was achieved by quantification of two DNA fragments following *Nla*III digestion.

The amount of 134 bp fragment, representative of the normally spliced *KIT* gene, and of 54 bp fragment, representative of the splice mutant *KIT*, were measured following electrophoresis using GeneScan software.

i. PCR to Produce DNA for Quantification

As described in example 2 section i. The reverse primer KIT35 is labelled with the ABI fluorescent dye FAM at the 5' end.

ii Restriction Enzyme Digestion

As described in example 2 section ii.

iii Electrophoresis and Quantification of DNA Fragments

Following digestion, 0.5 µl of the reaction volume was mixed with 2.5 µl of deionised formamide, 0.5 µl of GS350 DNA standard (ABI) and 0.4 µl blue dextran solution before being heated to 90°C for 2 minutes and rapidly cooled on ice. Three µl of this mix was then loaded onto a 377 ABI Prism sequencer and the DNA fragments separated on a 6% polyacrylamide gel in 1 X TBE buffer for 2 hours at 700 V, 40 mA, 32 W. The peak area of fragments representative to both the normal and splice mutant forms of *KIT* were quantitated using the GeneScan (ABI) software.

iv.Ratio Calculations

The peak area value of the 134 bp fragment (normal *KIT*) was divided by twice the peak area value of the 54 bp fragment (splice mutant *KIT*) in order to calculate the ratio value for each sample.

v. Results

Analysis was performed on animals from the Swedish wild pig/Large White intercross pedigree for which genotypes at *I* have been determined by conventional breeding experiments with linked markers. Figure 5 and Table 4 show the ratio of normal to mutant *KIT* calculated for animals from each of the three genotype classes, *I/I* (expected ratio 1:1), *I/i* (expected ratio 2:1) and I/I^P (expected ratio 3:1). The results are entirely consistent with the expected ratio values and indicate that the three genotype classes can be distinguished using this method.

TABLE 4

Ratio of the Two KIT Forms in Different Dominant White Genotypes in a Wild Pig/Large White Intercross

Genotype	Phenotype	Expected Ratio (Normal: Mutant)	Observed Ratio (Normal:Mutant) ± SE	Number Tested
I/I	White	1:1	1.15 ± 0.075	13
I/I^P	White	3:1	3.11 ± 0.084	12
I/i	White	2:1	2.23 ± 0.109	14

Figure 5 illustrates that the range of ratio values calculated for the two genotypes *I/I* and I/I^P do not overlap. This enables animals carrying the I^P allele to be identified and the frequency of the allele within different pig breeds determined. Ratio values were calculated for 56 Landrace and 33 Large White animals and the results are shown in Figure 6. A clearly bimodal distribution is

observed with 7 Landrace and 3 Large White individuals having a ratio value of approximately 3 or above, suggesting them to be heterozygous carriers for the I^P allele (genotype I/I^P). This means I^P has gene frequency estimates of 6.25% (7/112 chromosomes tested) and 4.5% (3/66 chromosomes tested) within the Landrace and Large White breeds respectively.

Example 5

Analysis for presence and quantification of the porcine KIT splice mutation using the PE ABI TaqMan chemistry

Method

i. Preparation of template DNA for PCR

DNA was prepared as in example 3, section i

ii. TaqMan® PCR reactions

TaqMan® PCR reactions were set up as shown in table 5

TABLE 5

PCR mix for TaqMan® based splice mutation test

Reagent	Final Conc[n]	Volume
10x TaqMan® Buffer A (Perkin Elmer)	1x	2.50 µl
25mM $MgCl^2$ Sol[n]	5mM	5.00 µl
DATP	200µM	0.50 µl
DCTP	200µM	0.50 µl
DGTP	200µM	0.50 µl
DUTP	200µM	0.50 µl
Amplitaq Gold™ (5U/µl) (Perkin Elmer)	0.05U/µl	0.25 µl
AmpErase™ N-Glycosylase (1U/µl) (Perkin Elmer)	0.01U/µl	0.25 µl
KITTM -NEST-F (5µM)	500nM	2.50 µl
KITTM-NEST-R (5µM)	500nM	2.50 µl
KITTM FAM (5µM)	100nM	0.50 µl
KITTM TET (5µM)	100nM	0.50 µl

25% Glycerol	8%	8.00 µl
Porcine genomic DNA		1.00 µl
		25.00 µl

The PCR primers used were as described below:

KITTM-Nest-F (5'-CTC CTT ACT CAT GGT CGA ATC ACA-3') and

KITTM -Nest-R (5'-CGG CTA AAA TGC ATG GTA TGG-3').

The TaqMan® probes used were:

KITTM-A FAM (5'-TCA AAG GAA ACA TGA GTA CCC ACG CTC-3') and

KITTM -G TET (5'- TCA AAG GAA ACG TGA GTA CCC ACG C -3')

The TaqMan® probes were prepared by Perkin Elmer and labelled with FAM and TET as indicated as well as the standard quenching group TAMRA. The 10x TaqMan® Buffer A, Amplitaq Gold™, AmpErase N-Glycosylase, NTP's and 25mM $MgCl_2$ used were part of the TaqMan® PCR Core reagent Kit, supplied by Perkin-Elmer.

The reactions were then placed into a Perkin Elmer ABI Prism 7700 Sequence Detector and the reaction carried out using the following thermal profile, 50°C for 2 minutes, 95°C for 10 minutes followed by 40 cycles of 95°C 15s, 62°C 60s. The reactions were carried out under the control of 'Sequence Detector V.1.6' software using the 'Single Reporter' and Real-Time' options with the 'Spectral Compensation' function activated. Upon completion of the run real-time profiles for each sample were examined on the ABI7700 to check for any samples giving highly irregular profiles which were then excluded. The thresholds for both dyes, Fam and Tet, were set so that they intercepted each

dye during the exponential phase of PCR. Following updating of the calculations in 'Sequence Detector V.1.6' software results were exportated into MS Excel for further analysis.

iii. Analysis of results

Based upon the underlying theoretical principle that one cycle of PCR gives a doubling in the amount of cleavage of the quenching dye from the allele specific probe and therefore doubles the signal one would expect the threshold cycle numbers from the II and Ii genotypes analysed to be as below:

Table 6:

Theoretical results for TaqMan® analysis of genotype at the KIT splice mutation

Genotype	Copies KIT 1 (G)	Copies KIT 2 (A)	Theoretical Ct TET (G)	Theoretical Ct FAM (A)
II	2	2	X	Y
Ii	2	1	X	Y+1

In theory the Ct for TET and FAM signals, represented as X and Y should be the same, as equal numbers of copies of the target sequences should be present in an *II* animal. However in practice this does not necessarily occur due to differences in the hybridization and cleavage efficiency of the two probes and variation in the setting of the threshold cycle between the two dye signals. The reduction in splice mutant containing (A) sequences relative to those <u>not</u> containing the splice mutation (G) in the *Ii* animals ie 2:1 G:A ratio rather than

1:1 as for *II* genotype, should lead to the FAM signal reaching the threshold 1 cycle later than the TET signal in the genotype *Ii* animals. The actual results for samples tested are shown in Table 7.

Table 7

Ct values from analysis of II and Ii genotypes

Sample	Genotype	Ct FAM (A)	Ct TET (G)	Ct FAM - Ct TET
1	*Ii*	24.68	22.59	2.09
2	*Ii*	25.98	23.62	2.36
3	*Ii*	26.54	25.57	0.97
4	*Ii*	27.37	24.78	2.59
5	*Ii*	24.94	21.61	3.33
6	*Ii*	25.68	22.1	3.58
				Ii Mean = 2.49
7	*II*	22.05	23.78	-1.73
8	*II*	24.22	24.59	-0.37
9	*II*	24.19	23.85	0.34
10	*II*	23.66	23.51	0.15
11	*II*	24.35	22.71	1.64
12	*II*	22.82	21.69	1.13
13	*II*	22.84	22.7	0.14
14	*II*	23.17	22.9	0.27
				Mean = 0.20

No Template	35	35	0
No Template	35	35	0
No Template	35	35	0
No Template	35	35	0

Despite variation around the mean values it can be seen from Table 7 that there is a significantly increased delay in the FAM signal reaching the threshold level (approximately 2 cycles) relative to the TET signal in *Ii* animals compared to *II* animals as predicted, reflecting the reduced number of copies of the splice mutant (A) sequence present in animals of the *Ii* genotype. Plotting of the individual samples on a scatter plot (Figure 7) shows clustering of the two genotypes with the *Ii* cluster shifted along the Ct FAM axis due to the reduced number of copies of the KIT2 (A) sequence for which the FAM probe is specific.

CLAIMS:

1. A method for determining coat colour genotype in a pig which comprises:

(a) obtaining a sample of pig nucleic acid; and

(b) analysing the nucleic acid obtained in (a) to determine whether a mutation is/is not present at one or more exon/intron splice sites of the *KIT* gene.

2. A method as claimed in claim 1 wherein the analysis in step (b) is carried out to determine whether a mutation is/is not present at the exon 17/intron 17 boundary.

3. A method as claimed in claim 2 wherein the mutation consists of the substitution of the G of the conserved GT pair for A.

4. A method as claimed in any one of claims 1 to 3 wherein the sample of nucleic acid is amplified prior to analysis.

5. A method as claimed in claim 4 wherein the nucleic acid is genomic DNA.

6. A method as claimed in claim 5 wherein amplification is carried out using PCR and at least one pair of suitable primers.

7. A method as claimed in claim 6 wherein the pair of suitable primers is:

5'-GTA TTC ACA GAG ACT TGG CGG C-3'); and

5'-AAA CCT GCA AGG AAA ATC CTT CAC GG-3'.

8. A method as claimed in any one of claims 5 to 7 wherein after amplification the nucleic acid is treated with a restriction enzyme, followed by analysis of fragment lengths.

9. A method as claimed in claim 8 wherein the nucleic acid is treated with the restriction enzyme *Nla*III.

10. A method as claimed in claim 8 or claim 9 wherein the ratio of restriction fragment lengths is determined.

11. A method as claimed in claim 4 wherein the nucleic acid is mRNA.

12. A method as claimed in claim 11 wherein the nucleic acid is amplified using RT-PCR.

13. A method as claimed in claim 12 wherein the length of RT-PCR product is determined.

14. A method for determining coat colour genotype in a pig which comprises the step of analysing a sample of pig KIT protein to determine whether the protein is the splice variant protein.

15. A kit for use in determining the coat colour genotype of a pig which comprises one or more reagents suitable for determining whether a mutation is present at one or more exon/intron splice sites of the *KIT* gene.

16. A kit as claimed in claim 15 which comprises one or more reagents for carrying out PCR and one or more pairs of suitable primers.

17. A kit as claimed in claim 16 which comprises the following pair of primers:

5'-GTA TTC ACA GAG ACT TGG CGG C-3'); and

5'-AAA CCT GCA AGG AAA ATC CTT CAC GG-3'.

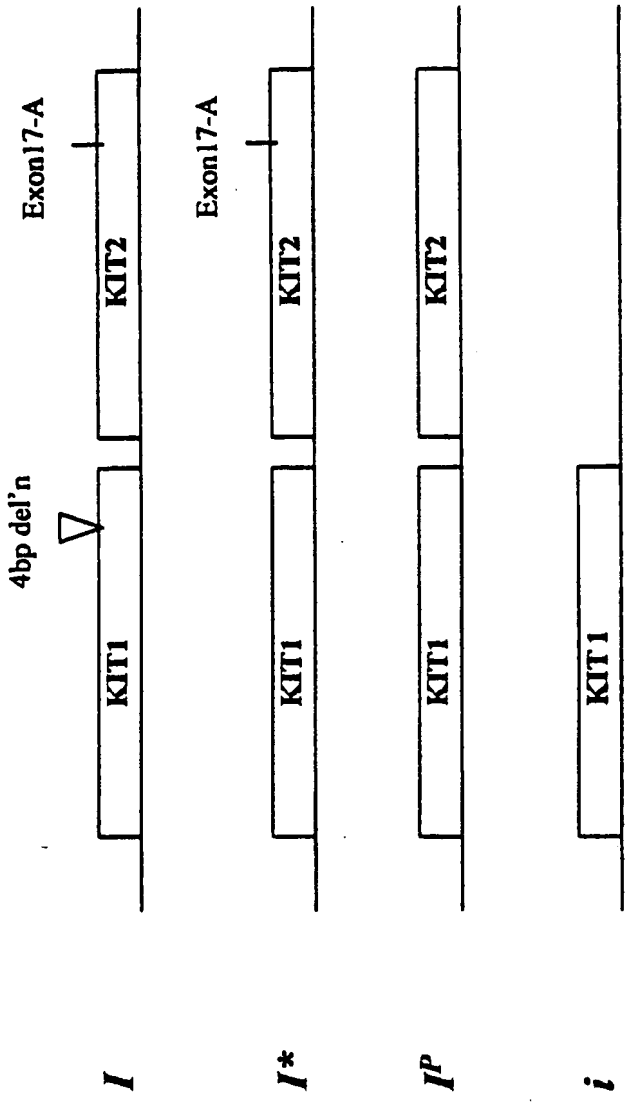

Exon17-A
KIT2
4bp del'n
KIT1
I
Exon17-A
KIT2
KIT1
I*
KIT2
KIT1
I^P
KIT1
i

FIG.1

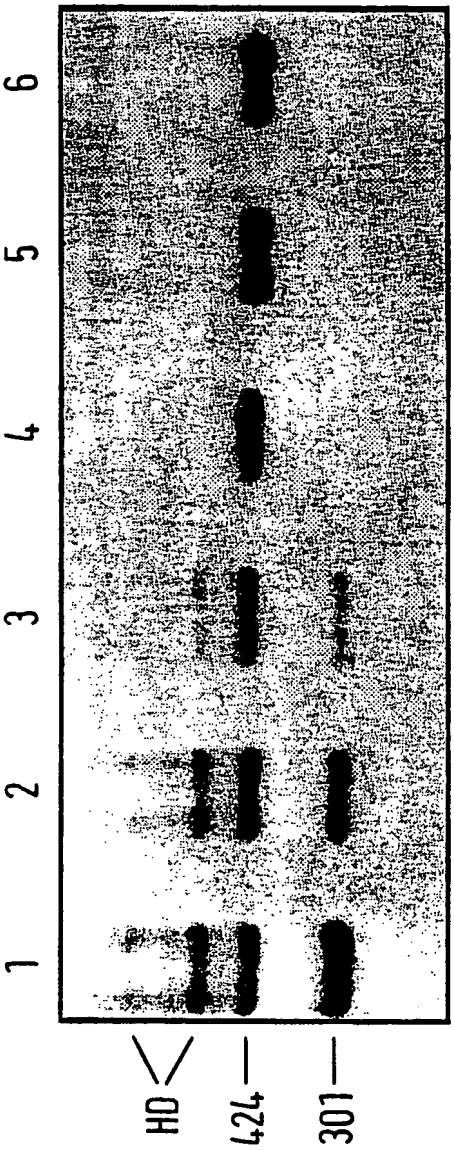


FIG.2

Sequence Alignment Across the Exon /Intron Border of KIT Exon 17

Allele	Gene Copy	Sequence (Exon 17 | Intron 17)
		AAT TAC GTG GTC AAA GGA AAC|GTG AGT ACC CAC GCT CTC CTG ACA GTC
I	KIT1	... ...|A.. ... ...
	KIT2	... ...|G... ... ...
I^P	KIT1	... ...|G.. ... ...
	KIT1	... ...|G.. ... ...
i	KIT1	... ...|G.. ... ...

FIG.3

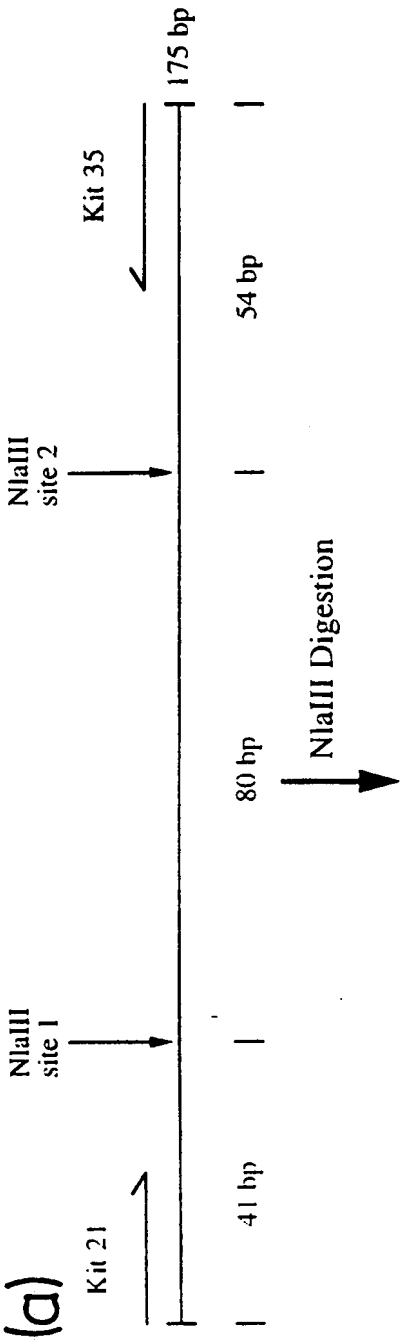

Kit 21
Kit 35
NlaIII site 1
NlaIII site 2
41 bp
80 bp
54 bp
175 bp
NlaIII Digestion

FIG.4(a)

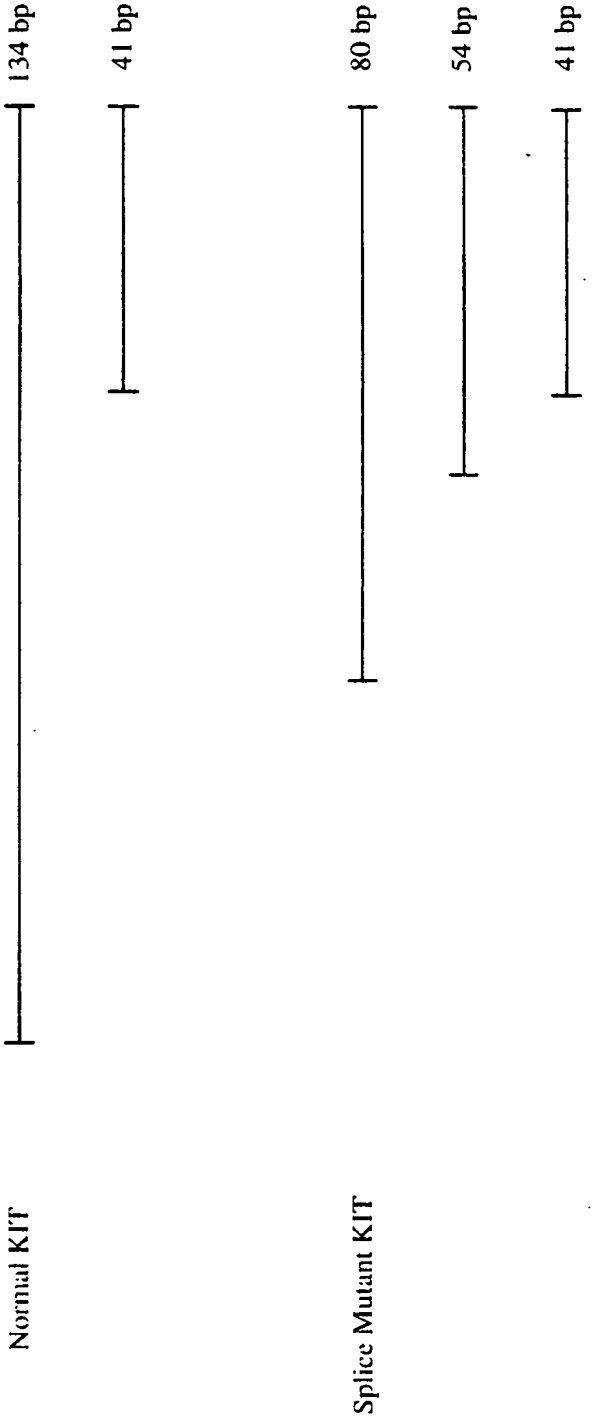

134 bp
41 bp
80 bp
54 bp
41 bp
Normal KIT
Splice Mutant KIT

FIG.4(b)

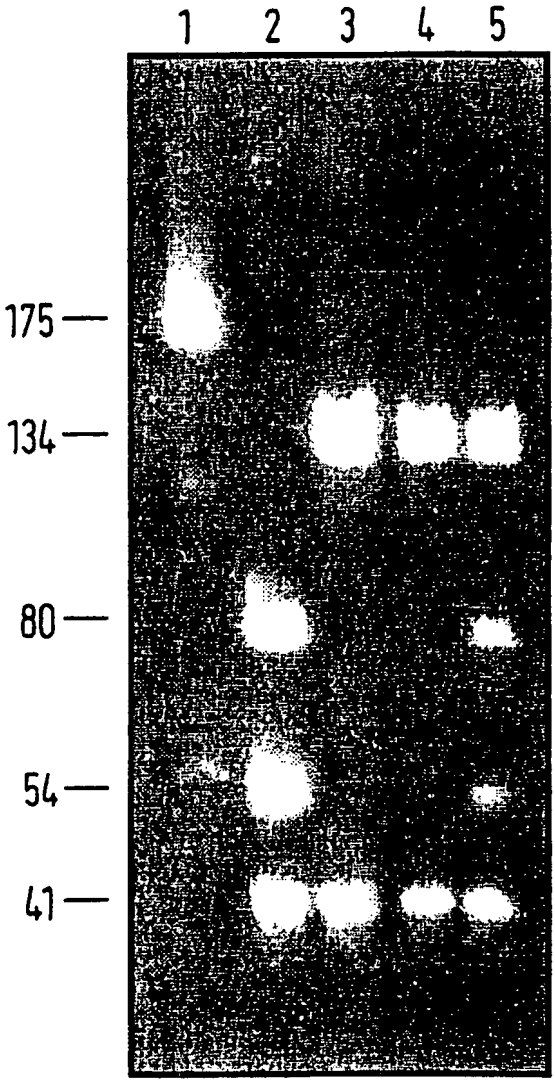


FIG.4(c)

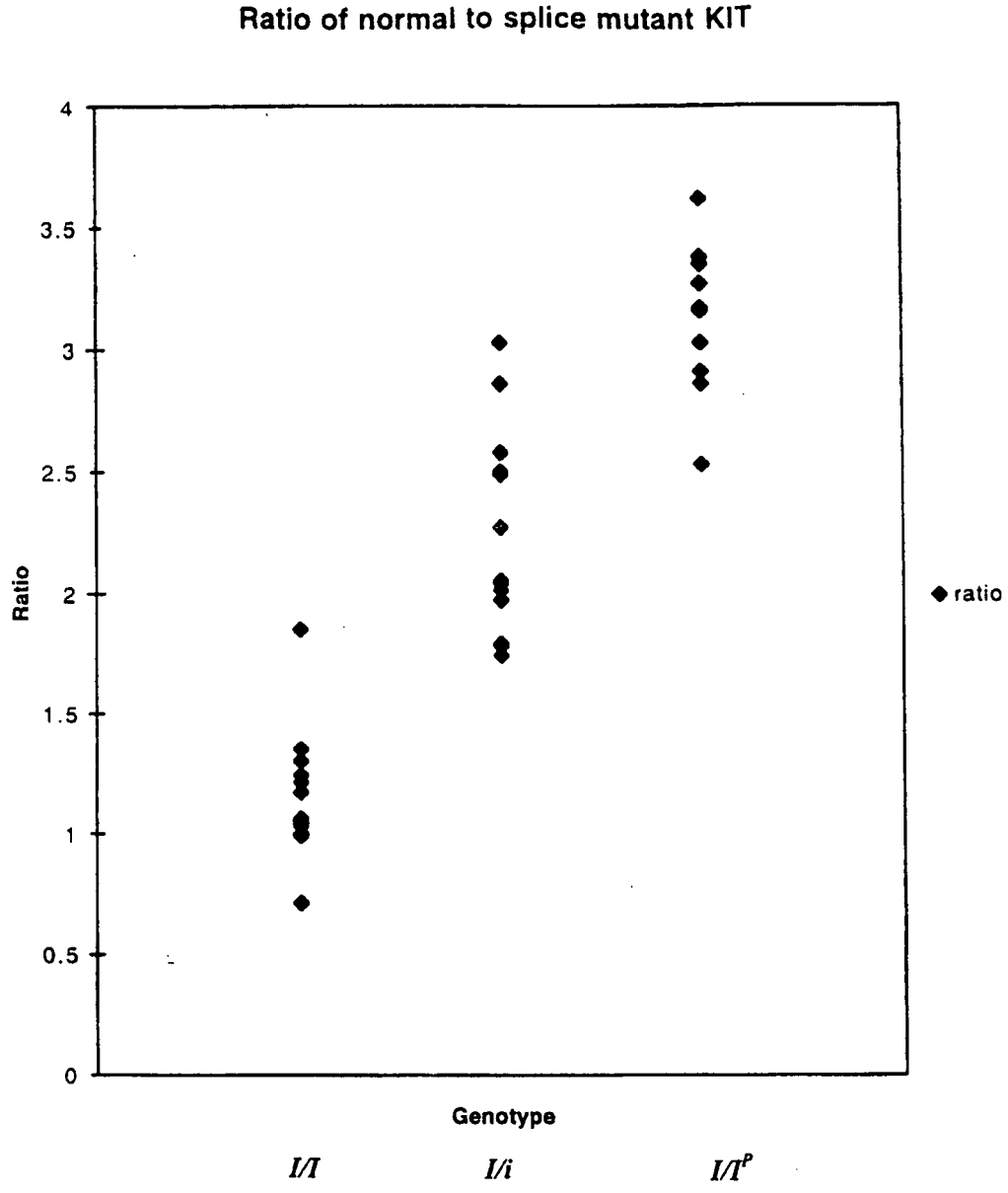

Ratio of normal to splice mutant KIT
4
3.5
3
2.5
2
1.5
1
0.5
0
Ratio
ratio
Genotype
I/I
I/i
I/I^P

FIG.5

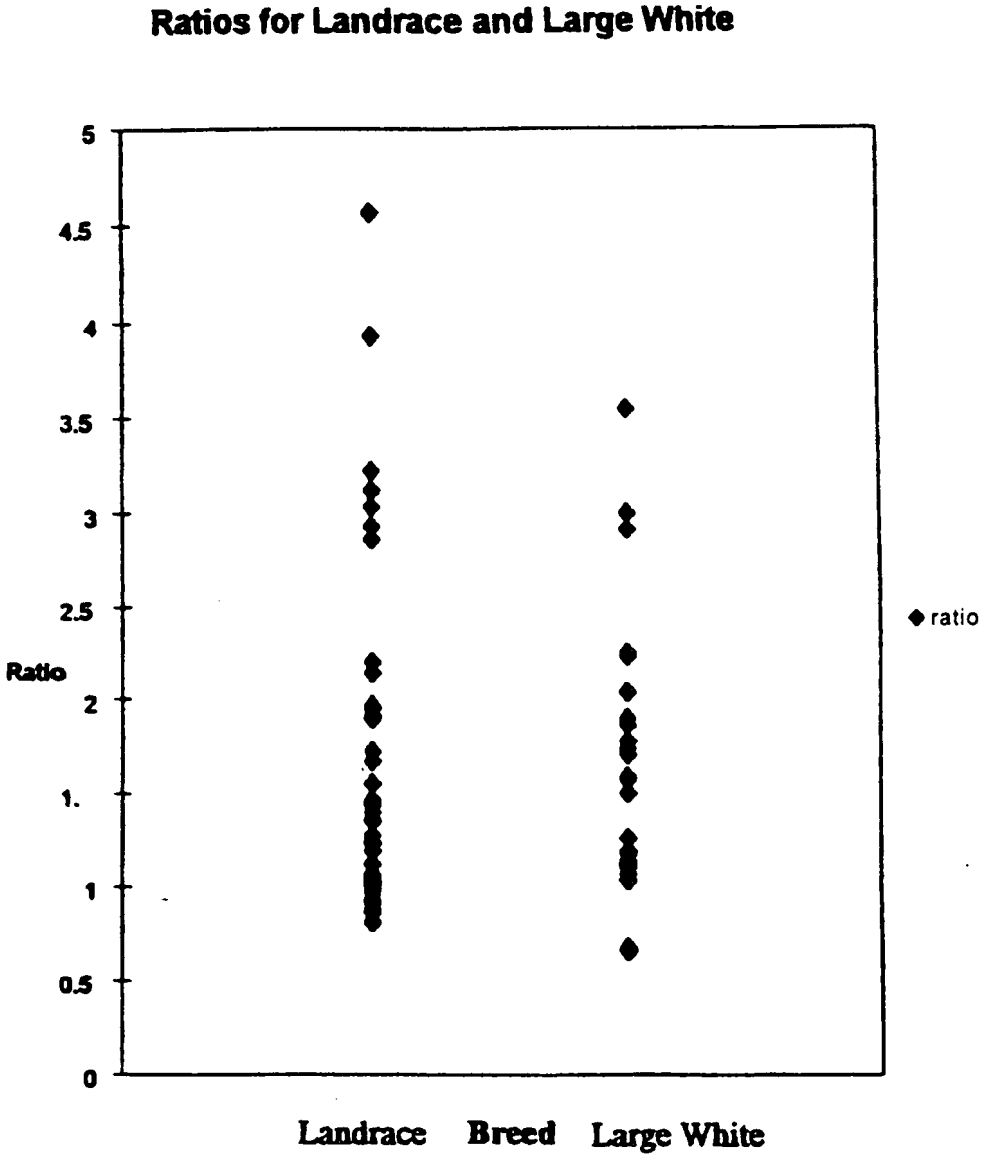

Ratios for Landrace and Large White
5
4.5
4
3.5
3
2.5
2
1.
1
0.5
0
Ratio
ratio
Landrace
Breed
Large White

FIG.6

Plot of Ct FAM vs Ct TET for TaqMan based PCR analysis of porcine KIT splice mutation genotype

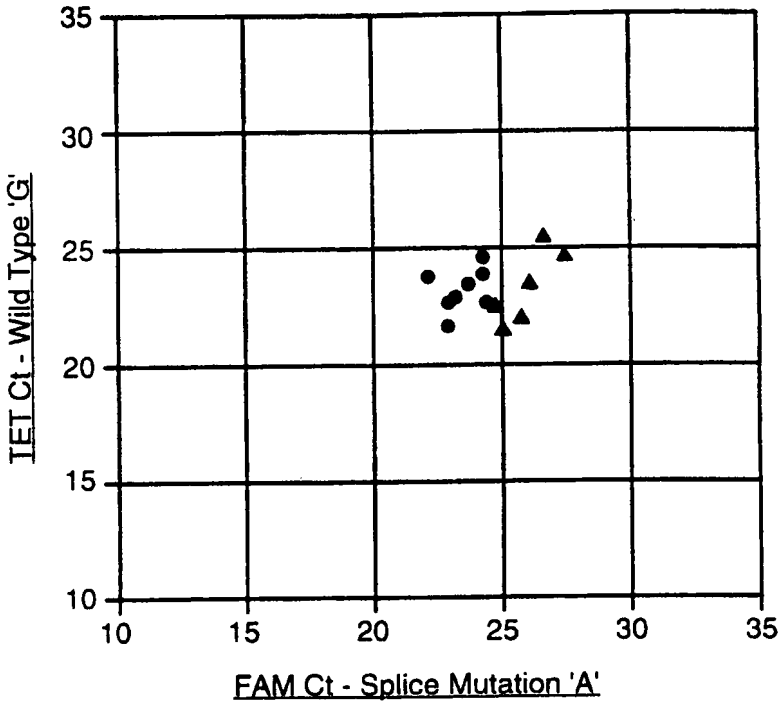


FIG.7

INTERNATIONAL SEARCH REPORT

Inte onal Application No
PCT/GB 98/03081

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12Q1/68 G01N33/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 97 05278 A (DALGETY PLC ;ANDERSSON LEIF (SE); MOLLER MARIA JOHANSSON (SE); WAL) 13 February 1997 cited in the application see the whole document ---	1,4-6
A	JOHANSSON MOLLER M ET AL: "Pigs with the dominant white color phenotype carry a duplication of the kit gene encoding the mast/stem cell growth factor receptor" MAMMALIAN GENOME, vol. 7, no. 11, November 1996, pages 822-30, XP002091653 see the whole document --- -/--	1,4-6

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
29 January 1999	16/02/1999

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Osborne, H

2

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Intv ional Application No
PCT/GB 98/03081

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	MARKLUND S ET AL: "Molecular basis for the dominant white phenotype in the domestic pig" GENOME RESEARCH, vol. 8, no. 8, 1998, pages 826-33, XP002091654 see the whole document	1-17

2

INTERNATIONAL SEARCH REPORT

.nformation on patent family members

Inte onal Application No
PCT/GB 98/03081

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9705278 A	13-02-1997	AU 6621296 A	26-02-1997
		CA 2227826 A	13-02-1997
		CN 1196755 A	21-10-1998
		EP 0842296 A	20-05-1998

JP009926

WO9849274.pdf
06/06/07 10:31 AM

WO9849274

Publication Title:

THERMOSTABLE DNA POLYMERASE AND INTEINS OF THE THERMOCOCCUS FUMICOLANS SPECIES

Abstract:

Abstract of WO9849274

The invention concerns a purified thermostable DNA polymerase, thermostable archae bacteria DNA polymerase of the Thermococcus fumicolans species having a molecular weight of the order of 89000 daltons and its thermostable inteins. Data supplied from the esp@cenet database - Worldwide

---

PCT ORGANISATION MONDIALE DE LA PROPRIETE INTELLECTUELLE
Bureau international

DEMANDE INTERNATIONALE PUBLIEE EN VERTU DU TRAITE DE COOPERATION EN MATIERE DE BREVETS (PCT)

(51) Classification internationale des brevets 6 : C12N 9/12 // 15/54

A1

(11) Numéro de publication internationale: **WO 98/49274**

(43) Date de publication internationale: 5 novembre 1998 (05.11.98)

(21) Numéro de la demande internationale: PCT/FR97/00761

(22) Date de dépôt international: 29 avril 1997 (29.04.97)

(71) Déposant *(pour tous les Etats désignés sauf US)*: APPLIGENE–ONCOR [FR/FR]; Parc d'Innovation, Rue Geiler de Kaysersberg, Boîte postale 72, F–67402 Illkirch (FR).

(72) Inventeurs; et

(75) Inventeurs/Déposants *(US seulement)*: QUERELLOU, Joël [FR/FR]; L'Arc'Hantel, F–29280 Brest (FR). CAMBON, Marie–Anne [FR/FR]; 8, allée de Stelle, F–29280 Plouzane (FR).

(74) Mandataire: BREESE–MAJEROWICZ; 3, avenue de l'Opéra, F–75001 Paris (FR).

(81) Etats désignés: CA, JP, US, brevet européen (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Publiée

Avec rapport de recherche internationale.

(54) Title: THERMOSTABLE DNA POLYMERASE AND INTEINS OF THE THERMOCOCCUS FUMICOLANS SPECIES

(54) Titre: ADN POLYMERASE THERMOSTABLE ET INTEINES DE L'ESPECE THERMOCOCCUS FUMICOLANS

(57) Abstract

The invention concerns a purified thermostable DNA polymerase, thermostable archae bacteria DNA polymerase of the *Thermococcus fumicolans* species having a molecular weight of the order of 89000 daltons and its thermostable inteins.

(57) Abrégé

La présente invention concerne une ADN polymérase purifiée thermostable ADN polymérase thermostable d'archaebactéries de l'espèce *Thermococcus fumicolans* ayant un poids moléculaire de l'ordre de 89 000 daltons, ainsi que ses intéines thermostables.

UNIQUEMENT A TITRE D'INFORMATION

Codes utilisés pour identifier les Etats parties au PCT, sur les pages de couverture des brochures publiant des demandes internationales en vertu du PCT.

AL	Albanie	**ES**	Espagne	**LS**	Lesotho	**SI**	Slovénie
AM	Arménie	**FI**	Finlande	**LT**	Lituanie	**SK**	Slovaquie
AT	Autriche	**FR**	France	**LU**	Luxembourg	**SN**	Sénégal
AU	Australie	**GA**	Gabon	**LV**	Lettonie	**SZ**	Swaziland
AZ	Azerbaïdjan	**GB**	Royaume-Uni	**MC**	Monaco	**TD**	Tchad
BA	Bosnie-Herzégovine	**GE**	Géorgie	**MD**	République de Moldova	**TG**	Togo
BB	Barbade	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tadjikistan
BE	Belgique	**GN**	Guinée	**MK**	Ex-République yougoslave de Macédoine	**TM**	Turkménistan
BF	Burkina Faso	**GR**	Grèce			**TR**	Turquie
BG	Bulgarie	**HU**	Hongrie	**ML**	Mali	**TT**	Trinité-et-Tobago
BJ	Bénin	**IE**	Irlande	**MN**	Mongolie	**UA**	Ukraine
BR	Brésil	**IL**	Israël	**MR**	Mauritanie	**UG**	Ouganda
BY	Bélarus	**IS**	Islande	**MW**	Malawi	**US**	Etats-Unis d'Amérique
CA	Canada	**IT**	Italie	**MX**	Mexique	**UZ**	Ouzbékistan
CF	République centrafricaine	**JP**	Japon	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Pays-Bas	**YU**	Yougoslavie
CH	Suisse	**KG**	Kirghizistan	**NO**	Norvège	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	République populaire démocratique de Corée	**NZ**	Nouvelle-Zélande		
CM	Cameroun			**PL**	Pologne		
CN	Chine	**KR**	République de Corée	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Roumanie		
CZ	République tchèque	**LC**	Sainte-Lucie	**RU**	Fédération de Russie		
DE	Allemagne	**LI**	Liechtenstein	**SD**	Soudan		
DK	Danemark	**LK**	Sri Lanka	**SE**	Suède		
EE	Estonie	**LR**	Libéria	**SG**	Singapour		

ADN POLYMERASE THERMOSTABLE ET INTEINES DE L'ESPECE THERMOCOCCUS FUMICOLANS

La présente invention concerne une nouvelle ADN polymérase thermostable ainsi que ses deux intéines, provenant d'une archaebactérie de l'espèce *Thermococcus fumicolans*.

Les ADN polymérases sont des enzymes impliquées dans la réplication et la réparation de l'ADN dans toute cellule vivante. On connaît aujourd'hui de nombreuses ADN polymérases isolées de micro-organismes tel que *E. coli* (ADN polymérase I) ou du phage T4. Des ADN polymérases ont aussi été identifiées et purifiées et à partir de micro-organismes thermophiles comme *Thermus aquaticus* (Taq polymérase, Chien, A. et *al.* J. Bact. 1976, 127:1550-1557 ; Kaladin et *al.* Biokhymiya 1980, 45:644-651), *Thermus thermophilus*, ou encore des espèces du genre *Bacillus* (demande de brevet Européen publiée sous le No. 699 760), *Thermococcus* (demande de brevet Européen No. 455 430), *Sulfolobus* et *Pyrococcus* (demande de brevet Européen publiée sous le No. 547 359). Parmi ces ADN polymérases issues d'archaebactéries on peut citer la *Pfu*, isolées de *Pyrococcus furiosus* (18), la Vent™ polymerase de *Thermococcus litoralis* (10), la 9°N de *Pyrococcus sp. 9°N* (15) et la DeepVent™ de *Pyrococcus GB-D*, les deux premières provenant de souches du littoral (Baie de Naples), les deux suivantes de souches sous-marines profondes.

Le mécanisme d'action des ADN polymérases est aujourd'hui relativement bien connu et consiste en une réplication de l'ADN à l'identique selon un mode semi-conservatif. Le brin recopié sert de matrice et les quatre nucléotides triphosphates sont le substrat de cette polymérisation. Les enzymes ayant une activité ADN polymérase sont aujourd'hui de plus en plus utilisées *in vitro* afin de travailler en biologie moléculaire dans divers buts tels le clonage, la détection d'erreurs, le

séquençage, le marquage, et de façon générale, l'amplification de séquences d'acides nucléiques.

Cette amplification, *in vitro*, de séquences d'acide désoxyribonucléique fait appel à la technique de la réaction de polymérisation en chaine (PCR) décrite dans les brevets Européens No. 200 362 et 201 184. Le principe de cette technique est basé sur la réalisation de cycles successifs d'extension d'amorces mettant en oeuvre les quatre nucléotides triphosphates ainsi qu'une ADN polymérase et un ADN matrice à recopier. A chaque cycle, l'enzyme double le nombre de brins d'ADN disponibles et entre chaque cycle une thermodénaturation est nécessaire afin d'ouvrir la double hélice d'ADN pour le cycle suivant. Les températures utilisées pour cette étape de thermodénaturation ne sont pas compatibles avec la conservation de l'activité de la plupart des ADN polymérases connues, telle la Klenow. C'est ainsi que des nombreux efforts de recherche ont été réalisés afin de trouver des enzymes supportant ces températures.

Cependant, si les micro-organismes thermophiles sont aujourd'hui connus, il reste encore difficile d'obtenir ces enzymes thermostables avec des rendements de production suffisants. La biologie moléculaire et le génie génétique permettent de palier cet inconvénient. Ainsi, une fois repéré dans le génome, le gène codant pour l'ADN polymérase est cloné, séquencé puis recloné dans un vecteur d'expression afin de produire la protéine dite alors recombinante, chez un hôte mésophile plus aisé à cultiver tel *E. coli* ou *S. cerevisiae*. Cette méthode d'expression chez *E. coli* a notamment été décrite dans la demande de brevet internationale PCT publiée sous le No. WO 89/06691 pour produire l'ADN polymerase de *Thermus aquaticus*.

L'ADN polymérase de l'invention provient d'une archaebactérie de l'espèce *Thermococcus fumicolans*. Outre ses propriétés thermostables la rendant particulièrement efficace notamment dans une processus de PCR, cette ADN

polymérase est remarquable en ce qu'elle présente deux "introns protéiques", encore appelés "intéines", au niveau de son polypeptide précurseur.

La séquence nucléotidique de ses intéines est insérée dans celle de l'ADN polymérase, généralement au niveau de sites conservés impliqués, après traduction, dans les réactions catalytiques. Ces séquences sont transcrites et traduites en même temps que celle de l'ADN polymérase et l'épissage autocatalytique des intéines produit alors trois enzymes: deux intéines et une ADN polymérase.

On trouve de telles intéines également au sein d'autres molécules telles l'ATPase vacuolaire chez *S. cerevisiae* (4), GyrA chez *Mycobacterium leprae* (7), Rec A chez *Mycobacterium tuberculosis* (5, 6). Les intéines font partie pour leur majorité de la famille des endonucléases de type "homing endonucleases" puisqu'elles coupent l'ADN en un site reconnu, à l'endroit même où leur séquence nucléotidique vient s'insérer.

Le développement des biotechnologies tant dans la recherche que dans les domaines de la médecine ou de l'agro-alimentaire, nécessite de disposer de divers types d'ADN polymérases susceptibles d'améliorer quantitativement et qualitativement des techniques aussi diverses que le clonage, la détection, l'amplification de séquences d'ADN. La présente invention vise précisément à offrir une nouvelle ADN polymérase thermostable qui est issue d'une espèce récemment décrite: *Thermococcus fumicolans* (8). Cet isolat a été isolé à partir de fragments de cheminées prélevées dans le bassin Nord-Fidgien lors de la campagne franco-japonnaise STARMER en 1989. Cette espèce, anaérobie stricte, présente une température optimale de croissance de 90°C, ce qui est relativement élevé pour un *Thermococcus*. Son pH optimum est de 8,8, et son taux de salinité de 20 g/l à 40 g/l.

L'invention a donc pour objet une ADN polymérase purifiée thermostable d'archaebacteries de

l'espèce *Thermococcus fumicolans* ayant un poids moléculaire de l'ordre de 89000 Da ainsi que ses intéines thermostables, dont le gène comportant les deux séquences codant pour lesdites intéines a été cloné.

Les travaux de recherche ayant permis d'identifier, de séquencer et d'étudier l'ADN polymérase de l'invention ont été réalisée à partir de la souche *Thermococcus fumicolans* ST557 déposée à la Collection de l'Institut Pasteur (CIP) sous le numéro CIP 104680. Cette ADN polymerase sera dénommée dans ce qui suit *Tfu*. Sa séquence de 774 acides aminés est représentée dans la liste de séquences en annexe sous le numéro SEQ ID NO:2. Un poids moléculaire de 89797 Da et un pI de 8.1 ont été déduits de cette séquence.

L'invention concerne donc l'ADN polymérase purifiée thermostable d'archaebactéries de l'espèce *Thermococcus fumicolans* ayant un poids moléculaire de l'ordre de 89 000 daltons ainsi que ses dérivés enzymatiquement équivalents. On entend par dérivés enzymatiquement équivalent, les polypeptides et protéines constitués par ou comprenant la séquence en acides aminés représentée dans la liste de séquences en annexe sous le numéro SEQ ID NO:2 dès lors qu'ils présentent les propriétés de l'ADN polymérase de *Thermococcus fumicolans*. A ce titre l'invention envisage plus particulièrement une ADN polymérase dont la séquence en acides aminés est représentée dans la liste de séquence en annexe sous le numéro SEQ ID NO:1 ou un fragment de celle-ci ou encore un assemblage de tels fragments, comme la séquence de 774 acides aminés représentée dans la liste de séquences en annexe sous le numéro SEQ ID NO:2.

En effet, la présence de deux intéines I-Tfu-1 et I-Tfu-2 dans la séquence numéro SEQ ID NO:1, sont susceptibles de conduire lors de la préparation par synthèse chimique ou par génie génétique, à des séquences d'ADN polymérase de *T. fumiculans* tronquées dont les

propriétés enzymatiques sont équivalente à celle de l'ADN polymérase de *T. fumicolans* purifiée.

On entend aussi par dérivés, les séquences en acides aminés ci-dessus modifiées par insertion et/ou délétion et/ou substitution d'un ou plusieurs aminoacides, pour autant que les propriétés de l'ADN polymérase de *T. fumicolans* qui en résultent ne soient pas significativement modifiées.

L'invention concerne également une séquence d'ADN constituée par ou comprenant la séquence codant pour une ADN polymérase de l'invention.

La séquence d'ADN représenté dans la liste de séquences en annexe sous le numéro SED ID NO: 1 représente une telle séquence. L'ADN codant pour l'ADN polymérase de *T. fumicolans* et ses deux intéines est constituée par le nucléotides 357 à 5028. Les nucléotides 1 à 356 correspondent à la région promotrice de ce gène. En conséquence, l'invention a pour objet une séquence d'ADN constituée par ou comprenant la séquence comprise entre les nucléotides 357 à 5028 de la SED ID NO: 1, ou un fragment de celle-ci ou encore un assemblage de tels fragments.

l'invention se rapporte tout particulièrement à une séquence d'ADN constituée par ou comprenant les nucléotides 357 à 1674 et 2755 à 3156 et 4324 à 5028 de la séquence d'ADN représenté dans la liste de séquences en annexe sous le numéro SED ID NO: 1.

Cette séquence code pour l'ADN polymérase de *T. fumicolans* dont la séquence de 774 acides aminés est représentée dans la liste de séquences en annexe sous le numéro SED ID NO: 2.

L'invention concerne autant l'ADN polymérase isolées et purifiées de la souche *Thermococcus fumicolans* que l'ADN polymérase préparées par synthèse chimique, par exemple par ligature de fragments polypeptidiques, ou encore par les méthodes du génie génétique.